

ON PSYCHOLOGICAL SIMILARITY
AND
PSYCHOLOGICAL DISTANCE.

A thesis presented to the Department of Psychology and
Sociology,
University of Canterbury.

In fulfillment of the requirements for the Degree of
Doctor of Philosophy.

by

William John Rosenberg.

December, 1975.

1975
BF
237
R 813
1975

ACKNOWLEDGEMENTS

To my supervisor, Professor R.A.M. Gregson,
for access to his mind, experience, papers and
patience.

To Sue Taylor, for access to her indubitable
typing skills.

To the technical staff and Dr Hugh Priest of the
Psychology Department, Professor J.J. Deely and Dr W.
Barit of the Mathematics Department, and the University
Computer Centre, for their hints, help, comments and
time.

To the New Zealand working taxpayer (through the
University Grants Committee) for three years
financial support.

To psychologists and others before me, without
whom this work would have been impossible.

CONTENTS

CHAPTER		PAGE
1.	INTRODUCTION	1
	1.1 Background.....	2
	1.2 Multidimensional scaling and similarity.	5
	1.3 Direct estimate models of similarity....	8
	1.4 Gross features of similarity and distance functions.....	9
	1.5 Discussion of these features.....	13
	1.6 Conclusions.....	24
	1.7 Notation and conventions used.....	31
2.	SURVEY OF RELEVANT LITERATURE.....	42
	2.1 Models of similarity.....	44
	2.1.1 Content models.....	46
	2.1.2 Distance models.....	66
	2.1.3 Other models.....	77
	2.1.4 Comparisons and conclusions.....	92
	2.2 Problems in multidimensional scaling	97
	2.3 Problems in psychophysics	102
	2.3.1 Ekman's Law.....	102
	2.3.2 Protheticness and metatheticness.....	104
	2.3.3 Category vs magnitude scales.....	105
3.	A PSYCHOLOGICAL DISTANCE FUNCTION.....	109
	3.1 The model on prothetic continua.....	112
	3.1.1 The model.....	112
	3.1.2 Discussion.....	120

CHAPTER		PAGE
3.	3.2 The model on metathetic continua.....	124
	3.2.1 The model.....	124
	3.2.2 Discussion.....	132
	3.3 Other comments.....	135
4.	PROPERTIES OF THE PSYCHOLOGICAL DISTANCE	
	FUNCTION.....	140
	4.1 The multidimensional case.....	140
	4.1.1 Properties discussed in Chapter 1.....	140
	4.1.2 Other properties.....	160
	4.1.3 D_{12} and D_{22}	168
	4.2 The one-dimensional case.....	173
	4.2.1 Incompatibility with Minkowski distance.	174
	4.2.2 The triangle inequalities.....	176
	4.2.3 Unidimensional similarity.....	178
	4.3 Summary	180
5.	SOME PROBLEMS OF PSYCHOPHYSICS FROM A SIMILARITY	
	VIEWPOINT.....	185
	5.1 Bisection, and category scales.....	187
	5.1.1 Bisection.....	187
	5.1.2 Category scales.....	198
	5.2 Other problems.....	206
	5.2.1 Helm's result.....	206
	5.2.2 Ekman's Law for metathetic continua.....	209
6.	APPLICATION TO SOME DATA AND TO MULTIDIMENSIONAL	
	SCALING.....	211
	6.1 Some similarity data.....	212
	6.1.1 In one dimension.....	212
	6.1.2 In two dimensions.....	226
	6.2 Multidimensional scaling.....	241

CHAPTER		PAGE
6.	6.2.1 Scaling with $D_{1\beta}$ or $D_{2\beta}$	241
	6.2.2 Monte Carlo studies.....	244
	6.2.3 Review of Eisler's experiment.....	261
	6.3 Some ordinal data.....	262
	6.4 Conclusions.....	265
7.	DISCUSSION AND CONCLUSIONS.....	268
	7.1 The modelling assumptions.....	270
	7.2 The models.....	273
	7.3 Applications to other areas.....	276
	7.4 Empirical evidence.....	278
	7.5 Multidimensional scaling.....	279
	7.6 General conclusions.....	280
	7.7 Suggestions for further study.....	282
	REFERENCES.....	285
	APPENDIX A.....	306
	APPENDIX B.....	314

LIST OF TABLES

TABLE		PAGE
1.1	Main notation used.....	36-38
2.1	Some vector Content models of similarity...	51
4.1	The triangle inequalities.....	158a
5.1	F-ratios and fitted parameters for two category scale models.....	202
6.1	Similarity data from Ekman (1965).....	214
6.2	Fit of best-fitting models for data in Ekman (1965).....	218
6.3	Similarity data from Eisler (1967).....	229
6.4	Data from dot stimulus experiment.....	235-236
6.5	Design for Monte Carlo studies.....	246
6.6	Results of Monte Carlo studies of POLYCON.....	254

LIST OF FIGURES

FIGURE		PAGE
1.1	Notation used in the vector interpretation of the Content models of similarity....	39
3.1	The equal-measure metathetic continuum.....	131
4.1	Isosimilarity contours for some Minkowski distances.....	161
4.2	Isosimilarity contours for some $D_{1\beta}$ and $D_{2\beta}$ distance functions.....	163-165
6.1	Fit and scaling of data from Ekman (1965), Expt. 1.....	219
6.2	Fit and scaling of data from Ekman (1965), Expt. 2.....	220
6.3	Fit and scaling of data from Ekman (1965), Expt. 3.....	221
6.4	Fit and scaling of data from Ekman (1965), Expt. 4.....	222
6.5	Eisler's experiment.....	227
6.6	Experimental design.....	232
6.7	City-block analysis of dot experiment.....	239
6.8	Euclidean analysis of dot experiment.....	240
6.9	Monte Carlo Study 2.....	248
6.10	Monte Carlo Study 3.....	249
6.11	Monte Carlo Study 5.....	250
6.12	Monte Carlo Study 8.....	251
6.13	Eisler configuration in terms of area and "shape".....	252

FIGURE

PAGE

6.14	POLYCON Monte Carlo Study: Euclidean analysis of D_{12} dissimilarities.....	256
6.15	POLYCON Monte Carlo Study: Euclidean analysis of D_{11} dissimilarities.....	257
6.16	POLYCON Monte Carlo Study: Euclidean analysis of D_{12} dissimilarities.....	258
6.17	POLYCON Monte Carlo Study: City-block analysis of D_{11} dissimilarities.....	259

ABSTRACT

A distinction is commonly drawn between "Content" and "Distance" models of similarity. The set theoretic interpretation of Content models is developed using Restle's (1959, 1961) structures. It is found that for spaces in which all dimensions are prothetic, a normalised type of model, similar to many vector content models, is correct: for $x, y \in \mathbb{R}^n$, $\beta > 0$,

$$D_{1\beta}(x, y) = \frac{(\sum |x_i - y_i|^\beta)^{1/\beta}}{(\sum |x_i|^\beta)^{1/\beta} + (\sum |y_i|^\beta)^{1/\beta}}$$

or

$$D_{2\beta}(x, y) = \frac{(\sum |x_i - y_i|^\beta)^{1/\beta}}{(\sum |x_i + y_i|^\beta)^{1/\beta}}$$

For spaces in which all dimensions are metathetic with no quantitative variation ("equal-measure metathetic") a Minkowski Distance model is correct. Different models are necessary for less simple cases.

The properties of the $D_{1\beta}$ and $D_{2\beta}$ distance functions are investigated. They have most of the invariance properties of the Content models and do not in general obey the triangle inequality or have additivity on straight lines. Their dissimilarity contours may be non-convex, and they disobey the additivity, subtractivity and decomposability properties set out by Beals, Krantz and Tversky.

Two similarity models of category scaling are suggested. Both predict the correct shape of the category-scale function for both prothetic and metathetic continua.

The similarity models are applied to some data of Ekman (1965); they give an excellent fit and correctly predict the nature of the respective continua. A Monte Carlo multidimensional scaling study using the $D_{1\beta}$ and $D_{2\beta}$ models showed that the normalised types of model are quite incompatible with the distance types of model: scaling of one type by the other gave low stress, highly interpretable solutions which were nonetheless quite invalid. An experiment of Eisler's is interpreted in the light of this result and of an auxiliary experiment.

It is concluded that, while the models have several weaknesses, they provide a rational basis, with prima facie empirical validity, for future similarity modelling. Two distinct types of model are identified - the normalised and the distance types - which are mathematically and psychologically incompatible. This demands a review of the validity of analyses of previous similarity experiments.

CHAPTER 1.INTRODUCTION

This dissertation is concerned with psychological similarity, psychological distance, and the unsatisfactory nature of some of the present mathematical models of these concepts. Modern psychology has increasingly recognised the concepts as important, even central, to the understanding or description of a wide range of psychological phenomena, including choice, stimulus generalisation and confusion, and discrimination. Probably reflecting this importance, direct similarity and dissimilarity judgments have been increasingly used, and a school built around various attempts to model these independently. Whether similarity in all these contexts is the same thing is a question for study, but from them, a rough picture of what psychological similarity and psychological distance are like seems to be emerging. To get a clearer understanding of what we are talking about however, it seems desirable to put this into historical context. We therefore give a very short and very rough sketch of the way the main models of similarity and dissimilarity have developed.

1.1 BACKGROUND

The early history of the use of mathematics in Psychology is largely the early history of psychophysical scaling. The early psychophysicists, from Fechner (who coined the term "psychophysics" - see, for example, Boring, 1950) onwards, were, like all psychophysicists, interested in the relationships between physical magnitudes and the corresponding sensations in the organism. They were, implicitly or explicitly, trying to find a function, hopefully of a general nature, relating the particular physical magnitude to its corresponding psychological sensation, and in doing so construct a scale for measuring sensations on that continuum.

The words "that continuum" indicate one of the two most obvious features which differentiate the earlier psychophysicists from their successors: unidimensionality. Both in the interests of simplicity and because of lack of suitable models, methods and computing techniques for analysing responses to multidimensional stimulus sets, early psychophysicists kept to the paradigm of holding all but one variable constant and studying each such variable individually. One of the major contributions of L.L. Thurstone (e.g. Thurstone, 1947) to psychophysics was to breach this barricade to the development of multidimensional

scaling. Thurstone not only gave a rational basis for models of most areas of psychophysics, but also developed the techniques of Factor Analysis to make use of judgments made on a multidimensional set of stimuli within the context of his models.

It should be noted though, that in his models, and in his use of Factor Analysis, Thurstone automatically assumed a particular way of combining the various dimensions or continua into a single quantity which underlay the observer's behaviour. This "combination rule" was the Euclidean distance: the distance of ordinary geometry, the distance of everyday life. It was chosen for its familiarity and mathematically pleasant nature rather than its plausibility as a model of any behaviour. In Factor Analysis, Thurstone neither used it explicitly (he used instead correlation coefficients which are closely related to Euclidean distance), nor, therefore, did he put it forward as representing a specific psychological construct, such as, for example, dissimilarity. It was left to Richardson (1938), his colleagues and following workers to begin the task of explicating what the distance function really meant.

The second feature that clearly differentiates the earlier psychophysics from the more modern

is its experimental methods. Under the (possibly correct) assumption that sensations are not directly measurable, virtually all judgments asked of observers were indirect ones: those of comparing differences and magnitudes, discriminating between magnitudes, detecting magnitudes. They were never asked how different or similar two magnitudes were, or how great a magnitude was. Starting with Fechner, and up to and including Thurstone, indirect methods were almost invariably used. The work of S.S. Stevens and his colleagues in the 1950's however, has brought about the recognition of direct judgment methods in psychophysics as a probably valid and almost certainly useful technique. They produced a specific model for psychophysics - that the sensation was a fixed power of the physical value on each continuum - and a distinction between prothetic (intensive) and metathetic (substitutive) continua, which appear to have somewhat different psychophysics (e.g. Stevens, 1957; Stevens and Galanter, 1957).

From the late 1950's the direct estimation techniques were applied extensively to similarity judgments (mainly by the Scandinavian school led by Gösta Ekman) with reliable results (e.g. Ekman and Sjöberg, 1965).

1.2 MULTIDIMENSIONAL SCALING AND SIMILARITY

Modern psychophysics thus has two major characteristics which distinguish it from both its beginnings and its not-too-distant past: the ability to deal with multidimensional stimuli, and the general acceptance and use of direct estimation methods.

Multidimensional psychophysics is, to a large extent, identified with the use of the powerful multidimensional scaling algorithms developed in the early 1960's by R.N. Shepard and J.B. Kruskal (Shepard, 1962a, b; Kruskal, 1964a, b). These are, in a sense, generalisations of the earlier factor analysis techniques as used (first by Richardson) to find the factors underlying the similarity or dissimilarity between stimuli, often found by direct similarity or dissimilarity estimation. Thus the modern multidimensional scaling algorithms were to at least some extent designed with the direct judgment technology in mind - but more of this later. However, one of their main contributions, leading largely from Shepard's (1957, 1958a, b) earlier work on stimulus generalisation, was that their structure made explicit some of the modelling assumptions necessary

in any multidimensional scaling technique. Firstly they recognised that the Euclidean distance was merely one of many possible distance functions. Its acceptance as a model of psychological distance was a matter for experimental verification just like any other model in psychology. Secondly, they made more explicit what the distance function meant psychologically and how it was to be used. The distance function was to represent, as Richardson had originally suggested, psychological distance. But whereas factor analysis assumed that psychological distance was a linear function of the distance function so that it could represent psychological distance and nothing else (unless fortuitously, or if the relationship to psychological distance was known), the new algorithms assumed only monotonicity between psychological distance and the distance function. Thus they could represent a wide range of similarity-related psychological phenomena - like probabilities of confusion, generalisation or discrimination, and of course similarity itself. As long as it was true that the behavioral measure in question was monotonic with dissimilarity, it could be represented by the distance function.

This of course puts great importance on the form of the distance function being used for the modelling. Although the class of monotonic functions is a very general one, there still remain important restrictions on the form of the distance function used. This is the burden of the theory put forward by Beals, Krantz and Tversky (e.g. Beals, Krantz and Tversky, 1968). That multidimensional scaling works using only "any monotonic function" is a warning of this fact: it is restrictive enough to give a quite rigid solution for a given distance function. It may of course be possible, as suggested before, that similarity has a different mathematical form in different contexts: there is already a substantial body of evidence to suggest this, and we will be investigating further in subsequent chapters. But the gross form of the similarity or distance function in each such context will need to have specific features which may be different from those in other contexts. The distance functions commonly used at present may not have all the required "gross features". What these "gross features" are, we will go into shortly.

1.3 DIRECT ESTIMATE MODELS OF SIMILARITY

When looking for a good model of dissimilarity or similarity, the obvious place to go is not, as most multidimensional scaling techniques have done, to a mathematical text on distance functions: mathematical tractability is a pleasant attribute, but should hardly be the critical one. The obvious place to go is in the area of the second major characteristic of modern psychophysics: direct magnitude estimation; and in particular, **direct similarity or dissimilarity estimation**. It seems strange that two major streams of modern psychophysical thought have barely interacted in this obviously desirable manner. The Scandinavian psychophysical school and its adherents have built up a great variety of information on direct estimates of similarity. They have proposed a great variety of models for these estimates, (see for example, Sjöberg, 1973; Lund 1974a), many of which seem ad hoc and poorly tested, but all of which fit data well enough to give their "gross features" enough credibility to at least rival those of models used in nonmetric multidimensional scaling. As will be seen, some of these features are inconsistent with the models of multidimensional scaling, yet on one

hand, the multidimensional scaling theorists and investigators have made no attempt to incorporate them into their algorithms, while on the other, the similarity estimation school appears to spend a large proportion of its effort trying to accommodate its results within the models of the multidimensional scaling algorithms. The cause may be technical problems (see Chapter 6) but it would seem worthwhile if such problems could be overcome.

1.4 GROSS FEATURES OF SIMILARITY AND DISTANCE FUNCTIONS

The models put forward by the direct similarity estimation school have been called "content" models (Ekman and Sjöberg, 1965), because of their set theoretic foundation, similarity being assumed to be the ratio of "common" to "total" properties. A great variety of such models have been put forward, all varying in their definitions of "common" and "total", and although they have almost always been tested in one special case, there are only three or four real contenders. We will not enumerate any of them at this point (see Chapter 2 and Lund, 1974a), but most have the following "gross features" that have some importance both psychologically and mathematically.

Noting that the content models are usually defined only for non-negative values, and using the notation of Table 1.1 (P. 36), with \mathbb{R}^n taken to represent psychological space, we have:

For every $x, y \in \mathbb{R}^{n+}$, $a \in \mathbb{R}^{n+}$, $\alpha \in \mathbb{R}$,

$$0 \leq \mathcal{S}(x, y) \leq 1. \quad (\text{S1})$$

$$\mathcal{S}(x, x) = 1. \quad (\text{S2})$$

$$\mathcal{S}(x, y) = 1 \Rightarrow x = y. \quad (\text{S3})$$

$$\mathcal{S}(x, y) = \mathcal{S}(y, x). \quad (\text{S4})$$

$$\mathcal{S}(x+a, y+a) > \mathcal{S}(x, y), \quad a \neq \mathbf{e}. \quad (\text{S5})^*$$

$$\mathcal{S}(\alpha x, \alpha y) = \mathcal{S}(x, y) \quad (\text{S6})$$

if $\alpha > 0$.

$$\mathcal{S}(x, y) = 0 \quad (\text{S7})$$

if x and y are on different axes.

$$\mathcal{S}(\mathbf{e}, x) = 0. \quad (\text{S8})$$

The "distance models" used in multidimensional scaling are generally the Minkowski distances given by

$$D(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^\alpha \right)^{1/\alpha}, \quad \alpha \geq 1,$$

*I have yet to find a rigorous proof of this condition for all the important content models, but it appears, via numerical examples, to be near enough for inclusion as "gross features".

although others, usually variations on this, have been used, and many others suggested (e.g. Young, 1972, Shepard, 1964). The "gross features" of these have been thoroughly investigated mathematically by Tversky, Krantz and Beals (Beals and Krantz, 1967; Beals, Krantz and Tversky, 1968; Tversky and Krantz, 1970), who have in fact derived necessary and sufficient conditions on the ordering of pairs of objects (x,y) for them to exist. They have also done some work on testing these ordinal conditions experimentally.

The "gross features" given here are somewhat different from those discussed by Beals, Krantz and Tversky, but they are chosen firstly because they correspond to those given for the content models of similarity, for which no Beals, Krantz and Tversky-type axiomatisation yet exists, and secondly for the psychologically useful and intuitive interpretations accessible from them. They are less general than those of Beals, Krantz and Tversky but some of their more important conditions can be derived as sufficient conditions for ours. Ours do not claim to supplant theirs however.

The features, corresponding to those given for the content models, are (using the same notation as before):

For every $x, y, z \in \mathbb{R}^n$, $a \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$

$$0 \leq D(x, y) < \infty \quad (D1)$$

$$D(x, x) = 0 \quad (D2)$$

$$D(x, y) = 0 \Rightarrow x = y \quad (D3)$$

$$D(x, y) = D(y, x) \quad (D4)$$

$$D(x+a, y+a) = D(x, y) \quad (D5)$$

$$D(\alpha x, \alpha y) = |\alpha| D(x, y) \quad (D6)$$

The following property has no direct counterpart in most of the content models, but is nonetheless a mathematically very important one:

$$D(x, y) \leq D(x, z) + D(z, y) \quad (D7)$$

A distance function with properties D1 - D4 and D7 is called a metric and should not be confused with the less general distance obeying all of the conditions D1 - D7.

Beals, Krantz and Tversky note the internal additive nature of these distances (interdimensional

additivity), the lack of interaction between dimensions (decomposability), and the fact that each dimension makes its contribution through a function of the difference between the values of the objects on that dimension (intradimensional subtractivity). If we let (rather loosely) $x_i = (0, 0, \dots, 0, x_i, 0, \dots, 0)$ for any $x = (x_1, x_2, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n)$, then a necessary condition for interdimensional additivity is

$$D(x, y) \geq D(x_i, y_i), \quad i = 1, \dots, n \quad (D8).$$

D5 is a necessary condition for intradimensional subtractivity and sufficient (as a special case) for a decomposability condition somewhat different from that of Beals, Krantz and Tversky.

1.5 DISCUSSION OF THESE FEATURES

1.5.1 Conditions 1 - 4

Looking at these conditions in more detail, it is clear that conditions 1 - 4 are more in the nature of boundary and rather trivial regularity conditions than anything basic in the structure. Perhaps the only really notable feature is that similarity is bounded, both above and below, while distance, at least defined on an unbounded space,

is only bounded below (condition 1). This will have a definite effect on the permissible forms of the transformation from one to the other. Since it is generally assumed that similarity (and distance) form a ratio scale, the absolute value of the upper bound in S1 is not necessarily fixed, but is usually taken as unity for convenience.

Conditions 2, 3 and 4 are usually merely assumed true: departures from 2 and 4 occur, but are not usually major enough to be significant. Conditions 3 of course, are not strictly true in all cases: if similarity is restricted by instruction or otherwise to only a subset of the attributes present (e.g. "Judge the similarity of the two objects with respect to their colour and size only", or, "when judging the similarity of these objects, ignore differences in colour"; both occur spontaneously in everyday life) then we may very well have two nonidentical objects rated as totally similar.

1.5.2 Condition 5

It is conditions 5 and 6 that are the critical ones. D5, S6 and D6 all take the form of functional equations which greatly limit the possible forms of ϕ and D . Conditions 5 relate to translations of the space. D5 states that

differences anywhere in the psychological space are equal. This, of course, has been a matter of great controversy, debate and investigation in psychophysics from Fechner to the present day. It is like saying that the psychological difference between having one candle and two candles (or between no candle and one candle for that matter!) lighting a room is the same as the difference between having one hundred and one hundred and one candles lighting the room, which is obviously false. Of course, one aim of psychophysics has been to construct a space in which equal distances appear equally different, so it could be said that D5 is acceptable if the right sensation scales are found. If the "right sensation scales" are, however, the direct estimation, power-law, scales of S.S. Stevens, as appears very likely, this would seem to be a reason for rejecting D5, at least in prothetic (intensive) continua.

There are three other points to note about Conditions 5. Firstly, leaving the space open to translation means that the origin has no fixed position. Thus it has no real interpretation in psychological terms. Conversely, the similarity space, with S5 holding has a fixed origin which can easily be interpreted, at least for intensive continua, as the point of zero intensity. The lack of this obvious and desirable property would

again seem to show a weakness in the distance condition D5. Secondly, the condition S5 gives some idea as to how the similarity function behaves (this is filled out further by S6). The further the objects are away from the origin (i.e. the greater the intensity on their constituent dimensions) the less psychological difference a given absolute change in the objects makes. This is consistent (qualitatively) with the candle example given above. Thirdly, as stated before, a decomposability condition is a necessary condition for D5. It states that the value on other dimensions has no effect on the difference between two values on any single dimension. This is weaker than Beals, Krantz and Tversky's decomposability condition (Tversky and Krantz, 1970), but it can be seen to be true in this case, and not for the content models, by putting

$$a = (a_1, a_2, \dots, a_{i-1}, 0, a_{i+1}, \dots, a_n), \text{ where } a \in \mathbb{R}^{n+},$$

in D5 and S5. Decomposability is an important condition in the theory of distances (see Tversky and Krantz, 1970) but has been disconfirmed in two of the three experiments we know of that test the condition specifically (Wender, 1971, and Krantz and Tversky, 1975).

Thus, certainly on the assumption of the power law of intensive continua, the "content" models would seem to fare the better of the two on conditions 5.

1.5.3 Condition 6

Condition S6 states the invariance condition for the similarity models. In a sense it is the opposite of D5, the invariance condition of the distance models. S6 states that all pairs of objects in the same proportion (in the sense that each respective dimension is in the same proportion for each pair of objects) have equal similarity. In each single dimension, this reduces to equal ratios being equally similar (or different). Thus the psychological difference between one and ten candles lighting a room would be the same as that between ten and one hundred candles. (Notice that nothing can be said about the psychological difference between no candles and any number of candles solely by S6).

On the other hand, D6 states that things twice as intense are twice as different: the difference between ten and a hundred candles is ten times that between one and ten candles. This, again is often absurd. This property is unfortunately usually glossed over in multidimensional scaling, however, because it is used to normalise the distances or configuration to keep their absolute

values within reasonable bounds.

Thus, again, with intensive continua, it would seem that the "content" models are far more plausible than the distance models with regard to conditions 6.

1.5.4 Conditions S7 and S8: the axes and origin

Feature S7 deals with objects on different axes. Once again, this points up the difference between the two classes of models in their interpretation of the axes: the distance models behave no differently on the axes than anywhere else in the space, while S7 is very reminiscent of the correlation coefficient with the axes as used in traditional factor analysis, where different axes were regarded as representing "totally different" or "independent" factors so that no similarity could be given between points on them. This interpretation of axes, while probably unnecessary and possibly undesirable, has been lost from multidimensional scaling. It would seem to imply that cross-modal matching would be difficult if not impossible.

S8 emphasises again the unique position of

the origin in the space. Here, everything is totally different from the origin: zero similarity exists. No such property exists for the distance models; on the contrary, it is a well-known (and much used - see for example, Shepard, 1964; Fischer and Micko, 1972; Hoijer, 1970b) mathematical fact that the properties of distances with properties D5 and D6 holding can be completely investigated by considering only distances to the origin. See, for example, Micko and Fischer, 1970, for further comment on this; it follows from the functional form of equations D5 and D6.*

*There has in fact been much misunderstanding over this point, with subsequent bad mathematical and methodological errors. Many authors confuse "metrics" - functions with properties D1 - D4 and D7 - with the much less general distances given by properties D1 - D7. Without D5 and D6, study of behavior around the origin and of isosimilarity contours (loci of points all at equal distance to the origin) is not sufficient, and the problem becomes much more complicated. Hoijer's work particularly is marred by this fault.

1.5.5 Condition D7: the triangle inequality

D7 is the so-called triangle inequality and is the most commonly quoted and blindly accepted condition on distances. It states that the distance on an indirect path between two objects is never shorter than the distance on a direct path. This makes sense if we can indeed add distances - i.e., think of distances as paths in the space. But this is not an obvious property of psychological space. In fact, the most usual interpretation of the triangle inequality is something like Shepard's (1962a, p.126) "if x is close to z and z is close to y, then x must be at least moderately close to y", which, as Shepard acknowledges, carefully avoids anything about adding measures of "closeness" or their opposites. Partly to overcome this, others have suggested such conditions as the even stronger "ultrametric" inequality (Johnson, 1967) (but see section 4.1)

$$D(x,y) \leq \max \{D(x,z), D(y,z)\}$$

which bypass the necessity for adding distances, but very little work seems to have been done on testing any such inequalities. In the case of the triangle inequality, this is probably largely due to uncertainty over what to use as a distance

measure and whether the properties of any such measure allow it to be added, but this is hardly a good reason for continuing blindly to accept its validity. In their work, Beals, Krantz and Tversky have in fact found ordinal conditions implying the triangle inequality so that in special cases it could perhaps be tested indirectly; but in general the triangle inequality is so deeply buried among other conditions of their axiomatisation that it would be difficult to sort out whether it was in fact the triangle inequality failing, if failures did occur.

Nonetheless, the interpretation of it given above makes it sound very plausible - even though there do exist exceptions: Shepard (1964) gives the example of "fable" being close to "table", and "table" being close to "chair", while "fable" is a long way from "chair". Clearly this is due to two spaces overlapping - the sound space putting "fable" close to "table", and the semantic space putting "table" close to "chair". Shepard suggests that the separation of the two spaces would be the answer to the problem. Thus, in some sense we would hope that while the triangle inequality may be violated sometimes,

it would hold "most of the time". Thus D7 should be taken less as a strict condition for a "reasonable" distance to comply with, and more only an indication of how we would like it to behave.

It might also be noted that some have suggested that similarity should comply with the condition

$$s(x,y) \geq s(x,z)s(y,z)$$

(e.g. Eisler - personal communication, 1974), an analogy to the triangle inequality due mainly to the common idea that similarity is an exponential decay function of distance. The same comments would apply to this as were made in reference to the triangle inequality - with the further remark that it is even less plausible, as to any psychological interpretation given to it of the form given to the triangle inequality.

A last comment on D7 concerns its conditions for equality: under what conditions do we have

$$D(x,y) = D(x,z) + D(z,y)?$$

For all the Minkowski distances, the answer is: when z lies between x and y on a straight line. For particular Minkowski distances, there are more general conditions for equality, but the "additivity on straight lines" condition is the most general and most important. It has widely been assumed to be desirable (see for example Beals, Krantz, and Tversky, 1968; Restle, 1959) but again seems to be attractive more mathematically than psychologically.

1.5.6 Condition D8

Finally, D8 expresses the idea that the overall difference between two objects is greater than the difference between them on a single salient dimension. This is easily defined in this case because of translational invariance (D5); the content models have a comparable property only when the difference on the single dimension is taken with certain restrictions on the values at which the other dimensions are held constant: The direction of the inequality depends on where in space the difference on the single dimension is taken. The basic concept seems to be an intuitively obvious one in some form and

has been tested and found to hold in one case using schematic faces (Tversky and Krantz, 1969). This test was, in fact, one of a contextually dependent version closer to that of the content models, and this fact serves as a warning as to possible counterexamples to the "pure" form of D8. These can occur when different dimensions are "weighted" differently in different situations. For example, when comparing cars, their colours are a relatively unimportant variable; thus two cars of the same model but quite different colours may be judged very similar - not very different. But the two colours when judged in isolation (i.e. with all other components zero) will be judged very different, so that D8 would apparently be violated.

1.6 CONCLUSIONS

As was noted previously, conditions 5 and 6 are the critical ones. They are critical on two counts. Firstly, they differentiate between the two models on the highly significant property of whether equal ratios of subjective intensities are subjectively equal, or equal differences of subjective intensities are subjectively equal.

Secondly, it can easily be seen that these two conditions along with S7 and S8, generate counterexamples to the hypothesis that the two types of model are monotone on each other.

On the first count, for intensive or prothetic continua, the content models are heavily favoured. However, while equal ratios may be approximately subjectively equal in prothetic continua, they may not be in metathetic or substitutive continua. Thus, the distance models may be favoured for metathetic continua. In Chapter 3 we will give a theoretical treatment that also suggests this. Nonetheless, for prothetic continua, models of the form of the content models would seem the best. A set-theoretic analysis, giving a more rational foundation to these models will be one of the main aims of this thesis.

The second count - non-monotonicity - has vital consequences for multidimensional scaling. If in fact a large range of stimuli could be best described by a model that was not monotone with any model presently used in multidimensional scaling (or even, perhaps, any function satisfying the triangle inequality - any metric), then the models for multidimensional scaling would have to be radically revised.

An example of non-monotonicity can be constructed quite simply: for any $x, y \in \mathbb{R}^{n+}$, we have, for any $\alpha \in \mathbb{R}$,

$$D(\alpha x, \alpha y) = |\alpha| D(x, y)$$

so that if $\alpha_0 > 1$ we have

$$D(\alpha_0 x, \alpha_0 y) = \alpha_0 D(x, y) > D(x, y),$$

and by S6,

$$g(\alpha_0 x, \alpha_0 y) = g(x, y).$$

Similarly, for any $a \in \mathbb{R}^{n+}$, $a \neq e$, we have by S5

$$g(x + a, y + a) > g(x, y)$$

and by D5

$$D(x + a, y + a) = D(x, y).$$

Now let $u = \alpha_0(x + a)$, $v = \alpha_0(y + a)$ for any such α_0 and a . Then

$$D(u, v) > D(x + a, y + a) = D(x, y)$$

and

$$g(u, v) = g(x + a, y + a) > g(x, y)$$

i.e., we have (u, v) and (x, y) such that

$$D(u, v) > D(x, y)$$

and $g(u, v) > g(x, y)$.

But, clearly, by choosing instead $0 < \alpha_0 < 1$, we can construct $u^1 = \alpha_0(x + a)$, $v^1 = \alpha_0(y + a)$ and

$$D(u^1, v^1) < D(x, y)$$

but $g(u^1, v^1) > g(x, y)$,

so that monotonicity is clearly violated.

Perhaps the major considerations here are the two related questions of the number of dimensions of the space and the identification of these dimensions. If the new models, when compared to the traditional ones, suggest in multidimensional scaling solutions that a different number of dimensions (preferably fewer) are being used, or that the dimensions can be given quite different, but equally plausible, interpretations, then previous analyses using the old models would have to be closely re-examined. If, on the other hand, it can be shown that the two types of models rarely give very different results, little is lost by continuing to use the more well-established procedures. Some evidence is already available on these points (Eisler, 1967; Roskam, 1972), and further will be given in this thesis.

Throughout this chapter it has almost always been assumed that objects serving as stimuli could be described as points in n -dimensional real space. This is probably not true for all stimuli. For this reason "nondimensional" models and corresponding computing algorithms have been developed to analyse stimuli in other ways. Most of these use clustering techniques

of various kinds. These cluster together very similar stimuli, and clusters are interpreted according to the common properties of their constituents (see for example, Johnson, 1967; Arabie and Shepard, 1973; Arabie and Boorman, 1973; Boorman and Arabie, 1972). They all of course assume some sort of distance or similarity function to determine what is "very similar", but we will not be going into these in any detail. This is not to imply anything about the relative usefulness of dimensional versus clustering techniques: both clearly have their place. Hopefully though, the models developed here throw some light on clustering techniques also.

Cunningham and Shepard (1973) have also begun the development of a new non-dimensional type of algorithm: "maximum variance non-dimensional scaling", which is mid-way between clustering and multidimensional scaling. It is a further generalisation of non-metric multidimensional scaling, now allowing any monotone function of any metric, instead of any Minkowski (or other specified) distance. Since it does not give a mapping of points in multidimensional space, its main importance seems

to be in finding the form of the generalisation function although it may also be of some use in conjunction with other scaling techniques. One of its central features is its insistence on the maximal satisfaction of the triangle inequality - this is one of the conditions used by the algorithm to find a solution. In view of the comments made in the previous section, the validity of this is questionable: the assumptions involved may again turn out to be more rigorous than thought, and in the wrong direction. In particular, it is difficult to see what is hidden beneath the maximum variance goal: it may, for example, force many triangle equalities, with subsequent effects on the possible forms of the metric. Though again, our investigation may shed some light on these points, we will not be going into it in greater detail either.

However, all these scaling techniques are relevant to the present thesis in the following way: one of our basic propositions is that the distance functions used in scaling techniques are all-too-often assumed general enough to cover most ranges of stimuli. Sometimes they are used without even considering in detail their specific

form and what it implies. Here, we are trying to spell out some of these details, indicating where they may be wrong, and suggesting a more plausible (but likely still incorrect) alternative. We want to make specific that even non-metric multidimensional scaling involves modelling of a very definite nature, logically little different from say, the (albeit better specified) models of learning theory.

This thesis will, then, in the main, be investigating two closely related topics:

- (a) the adequacy of present distance functions for modelling psychological similarity, either in the form of direct judgments or as implicit in other behaviour; and
- (b) the adequacy of present distance functions for use in psychophysical scaling.

In the process of this investigation, other related areas will naturally arise, including the form of the "generalisation gradient", or the relationship between similarity and distance; the place of the prothetic-metathetic continuum distinction in similarity judgments; and the general implications

of the similarity models developed to scaling theory.

Chapter 2 will give a survey of work done to date on related subjects. Chapter 3 will give a theoretical development of a content model and a related distance function with many of the desirable properties discussed above. In Chapter 4 some of these properties will be investigated to give a better picture of how the model behaves, while Chapter 5 will attempt to show some of the implications for psychophysical scaling. Data, both artificial and empirical, will be considered in Chapter 6, and the implications for multidimensional scaling discussed. Finally, Chapter 7 will present conclusions and suggestions for further work.

1.7 NOTATION AND CONVENTIONS USED.

In the discussion above, and in the following chapters, several terms will occur repeatedly: model; similarity, dissimilarity, difference, distance; dimension, attribute, property, characteristic, factor, quality, feature, aspect. We will try here to clarify their meanings as much as seems necessary.

A mathematical "model" of some behavior is generally thought of as something that explains

that behavior in that it reduces that behavior to simpler or more easily understood processes. In this sense, it constructs the observed behavior from more "basic" processes. Thus, for example, a paired-associate learning model might reduce this type of learning to a series of probabilities of conditioning of stimulus elements (Atkinson, Bower and Crothers, 1965, p.85). This is the strongest and best sense of the word "model".

It is often used, however to name any mathematical construct which is claimed to no more than describe the behavior in question. It no more "explains" the behavior than a physical description of a car "explains" how the car works. Certain things can be predicted from it, if it is a good description, and certain of its primitives will perhaps be interpretable in the light of other, present or future, knowledge (though by no means necessarily).

This is the sense in which "model" is generally used in reference to psychological similarity or difference. To say that a certain Minkowski distance is a "model of psychological dissimilarity" does not mean that we have any deeper understanding of the mechanisms of similarity judgments: it only

means that we have a more or less accurate way of summarising - describing - its behaviour in a short mathematical formula. Thus, most of the "models" of similarity presented here do not and should not claim to "explain" anything: they are, at most, purely descriptive, and any "interpretation" given to their properties is post hoc. "Model" will be used here in this, its weakest, sense - to mean no more than a description of the behavior concerned - unless otherwise explained. It is used not because we feel it is a good term in the context, but because it has been used too often to be easily dropped.

The concepts of psychological "similarity", "dissimilarity", "difference" and "distance" will not be defined. They are to be taken as primitives without further explanation, except in the case of their corresponding direct judgments, when they will be defined operationally as the (numerical) results of these judgments. Except when drawing distinctions between them (as, for example, when discussing the relationship between similarity and dissimilarity), they will be used interchangeably and will be assumed to refer to the same psychological process unless otherwise stated. This is not necessarily to imply that in all contexts

in which similarity occurs, similarity has the same form - it probably does not - but within each such context the four (or other, equivalent) terms will be taken as explained.

On the mathematical level, confusion often occurs over the use of the words "distance" and "metric". A metric is usually strictly defined as a real-valued function of two variables (not necessarily in \mathbb{R}^n) satisfying conditions D1 - D4 and D7 of section 1.4 above, (e.g. Kelley, 1969, p.118). A distance is sometimes confusingly defined to be a metric, but at other times is used in other ways. For example, Beckenbach and Bellman (1961, p.99-100) define a distance to be one conforming to conditions D1 - D7, while its use in the theory of convex functions and convex sets is very similar (see for example, Eggleston, 1958, p.54ff.). We will use the following terminology: a "metric" will be defined as above; a "distance"* will be a function satisfying D1-D7; and a "distance function" will be any function that generally measures "apartness" in the space in a way analogous to Shepard's (1962a, p.126) concept of "proximity measure". A "distance function" - as distinct from a "distance" - will not necessarily

* Our "distance" corresponds to what is also sometimes called a Minkowski metric (see Busemann, 1955, p. 94, p. 100; Micko and Fischer, 1970).

have any one of the properties D1 - D8. The distinction between these three terms must be always born in mind, as otherwise bad mathematical errors can be (and have been) made.

The words "dimension", "attribute", "property", "characteristic", etc. are probably the set of most often confused words in the area we are working in. One of the clearest definitions of them is given by Torgerson (1958, p.248-249). We will be roughly adhering to Torgerson's conventions, but without being too rigorous when there seems little likelihood of confusion. In particular, we may sometimes use "attribute" and "dimension" interchangeably, and, more often, but in a different context, "attribute", and "property", "factor", "characteristic", "quality", "feature", or "aspect".

Finally some notation. We will list it first and then give a few words of explanation.

TABLE 1.1MAIN NOTATION USED

<u>SYMBOLS</u>	<u>DESCRIPTION</u>
Lower case latin alphabet (a,b,c,...x,y,z)	Stimulus names; the values of the stimuli in one-dimensional space.
Upper case latin alphabet (A,B,C,...X,Y,Z)	The sets of aspects corresponding to the stimuli a,b,c,...x,y,z, respectively.
(x_1, x_2, \dots, x_n)	The vector or point in n-dimensional real space, representing stimulus x, assuming such a representation exists.
x_i	The set of aspects in the ith attribute of stimulus x. $x = \bigcup_{i=1}^m x_i, \quad x_i \cap x_j = \emptyset, i \neq j.$
h_x (*)	$(\sum_{i=1}^n x_i ^2)^{1/2}$, the length of vector (x_1, x_2, \dots, x_n) in Euclidean space.
ϕ_{xy} (*)	The angle between vectors (x_1, x_2, \dots, x_n) and (y_1, y_2, \dots, y_n) .
p_{xy} (*)	$h_x \cos \phi_{xy}$, the length of the projection on (y_1, y_2, \dots, y_n) of (x_1, x_2, \dots, x_n) .

TABLE 1.1 continued

<u>SYMBOLS</u>	<u>DESCRIPTION</u>
m	Some measure (to be defined in the context in which it is used) on the space of all aspects.
$\beta(x,y) \quad (+)$	The psychological similarity between stimuli x and y .
$D(x,y) \quad (+)$	The psychological dissimilarity between stimuli x and y .
\mathbb{R}	The real line.
\mathbb{R}^+	The non-negative real line.
\mathbb{R}^n	n -dimensional real space (a space of either vectors or points).
\mathbb{R}^{n+}	Those vectors or points in \mathbb{R}^n with all components non-negative.
Lower case Greek alphabet ($\alpha, \beta, \gamma, \dots$)	Real numbers ("scalars").
$\cup, \cap, \subset, \subseteq, \supset, \supseteq, -, \varepsilon, \Delta$.	The usual notation of set theory; $A \Delta B = (A - B) \cup (B - A)$.
$\Rightarrow, \forall, \exists, \ni$.	Respectively: implies; for all; there exists, or, for some; such that.
\Leftrightarrow , iff	If and only if.

TABLE 1.1 continued

<u>SYMBOL</u>	<u>DESCRIPTION</u>
\emptyset	The empty set
$x+y$	$(x_1+y_1, x_2+y_2, \dots, x_n+y_n)$ (where $x, y \in \mathbb{R}^n$)
αx	$(\alpha x_1, \alpha x_2, \dots, \alpha x_n)$ (where $x \in \mathbb{R}^n$)
e	$(0, 0, \dots, 0)$ - the origin
ψ	The psychophysical function

FOOTNOTES:

(*) See Figure 1.1: "Notation used in the Vector Interpretation of the Content Models of Similarity".

(+) Various models for these will be defined
- see Chapters 2 and 3.

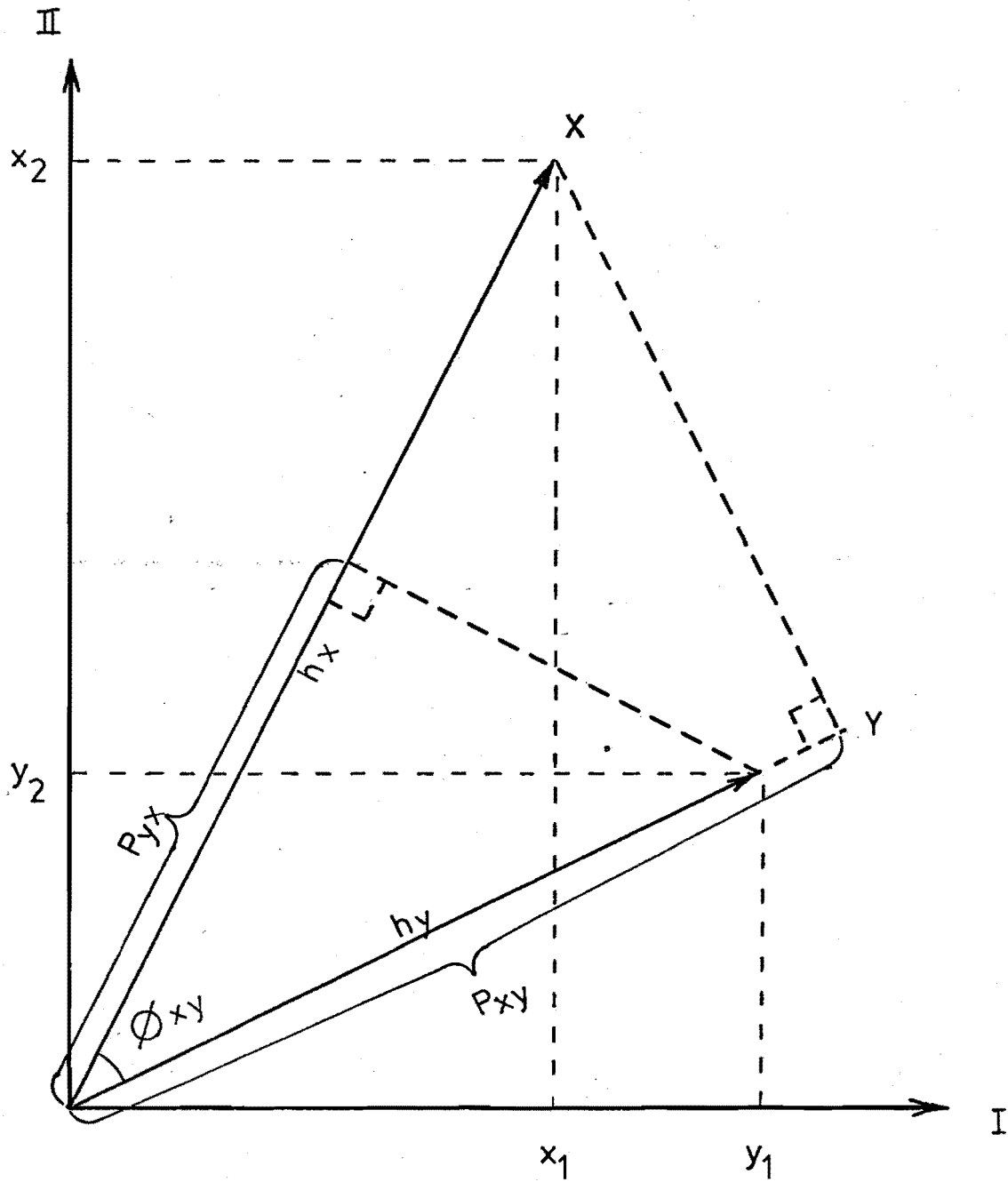


FIGURE 1.1 Notation used in the Vector interpretation of the Content Models of Similarity. To be generalised to n dimensions.

Although a formal distinction is drawn here between stimulus names and their representations in either aspect space or multidimensional real space, this will not be strictly adhered to. A stimulus x will sometimes be identified with its set of aspects, X , or with its vector or point representation, (x_1, x_2, \dots, x_n) in psychological space. A more detailed and rigorous explanation than that used here, of the representation of a stimulus in terms of its aspects, can be found, for example, in Tversky (1972).

$\mathcal{S}(x, y)$ and $\mathcal{D}(x, y)$ will represent actual models or judgments of psychological similarity or dissimilarity respectively. When an abstract model behaving rather like similarity or dissimilarity (distance, difference, etc.) is being referred to, the notation $s(x, y)$ and $d(x, y)$ will be used respectively. Usually however we shall use $\mathcal{S}(x, y)$ and $\mathcal{D}(x, y)$, implying a modelling process.

Finally, a note on the space used. In writing \mathbb{R}^n , we do not necessarily imply that we are referring to an n -dimensional physical space; in fact we will usually be referring to n -dimensional psychological space. We will make it clear in the

context what space we are working in, but if for example we are in psychological space and refer to it as \mathbb{R}^n then we are making the assumption that the psychological space can be represented by \mathbb{R}^n .

CHAPTER 2.

SURVEY OF RELEVANT LITERATURE.

Three main areas will be traversed in the following chapters: different models of similarity; some of controversies in psychophysics; and the use of multidimensional scaling algorithms. The latter two areas will be approached only from the context of the first, which is the core of the thesis. In this chapter we shall therefore first look at some of the many and varied formulas that have been suggested as suitable representations of similarity. Ekman and Sjöberg (1965) introduced a major distinction between two classes of similarity model: the "Content" models and the "Distance" models. Both classes have many competing members but they are not exhaustive of all models that have been put forward. Other areas of experimental and theoretical psychology have produced various other models which either fall between, or cannot be directly related to, the two classes. A third class of similarity model comes from the use of purely arbitrary "similarity coefficients", mainly in taxonomic work. Some work has further been done on comparing models - mainly in the Content versus Distance model arena.

For more general surveys of similarity models, Gregson (1975) is the most detailed, but Luce, Bush and Galanter (1963), especially chapters 2 - 5, also summarise some very useful material.

The main controversies we shall be entering into in psychophysics will be concentrated around the questions of the form of the psychophysical function for category scales, and its relation to that for direct estimation scales. This will also involve discussion of the Metathetic (qualitative) versus Prothetic (quantitative) continuum distinction, which will also take an important place in the central developments of Chapter 3.

Because of the interrelationship between similarity models and multidimensional scaling (MDS) (see Chapter 1), we shall be considering how our models affect MDS results. We are thus interested in the artefactual and misleading solutions that such algorithms tend to produce, but less interested in their technical details. Shepard, Romney and Nerlove (1972, Volumes I and II) give a good detailed survey of this area.

2.1 MODELS OF SIMILARITY

Similarity has long been recognised as a fundamental concept in psychology. For example, Mach, (1914, Chapter 14, sections 4-11) discussed the invariances of visual stimuli that lead to "sameness of form". It took a central place in the Gestaltist view of psychology, the "Factor of Similarity" being one of Wertheimer's Laws of Organisation in Perceptual Forms (Wertheimer, 1938). This used similarity as a primitive concept in describing the more complex ideas of Form; others working largely in the Gestaltist tradition investigated similarity in some depth. For example, Ternus (1938) looked at identity of features in pairs of objects, while Goldmeier (1972, 1936) carried out a major experimental study of similarity.

Research of a largely qualitative and discursive nature is still widespread: for examples see Derks (1972); Bindra, Donderi and Nishisato (1968); Levy and Kaufman (1973); Lockhead (1970); Wallach (1958). It supplies basic data that will have to be explained (and largely has not been) by quantitative models of similarity. The effects of singularity (or prägnanz) and context (set) are

shown by these authors (e.g. Wertheimer, Goldmeier, Wallach, op.cit.) to be critical ones in similarity behaviour, but they have yet to be given a general explanation by any current mathematical model. They could form the basis of a system of axioms from which a similarity model would be constructed; more likely, they will serve to highlight lines of demarcation between areas where different models of similarity may apply.

The use of quantitative models of similarity tightens the structure of any theorising. It makes the ad hoc nature of many theories more obvious and clarifies differences in similarity behaviour in different contexts. It also makes testing more difficult because false disconfirmation of a theory becomes much easier; and specification of primitives becomes more objective and hence, at times, quite difficult. Some of the problems in this area have been considered by Fillenbaum (1973): choice of stimulus sample, wording of instructions, restrictions on responses, interpretation of results, and the interactions between all these. However, the quantitative models are a definite step forward towards making the theories of psychology part of a rigorously testable science.

2.1.1 CONTENT MODELS

2.1.1.1 Models

A large variety of Content models have been suggested, and tested to varying extents. Sjöberg (1973) outlines the development, strengths and weaknesses of many of them, so we shall often refer to him where relevant.

The first ad hoc model of similarity seems to be that of Eisler and Ekman (1959), put forward for unidimensional similarity judgments. After trying various other formulas, they found that

$$s(x,y) = \frac{\min(x,y)}{(x+y)/2} \quad (2.1)$$

fitted the data well when x and y were found by fractionation procedures in the continuum of pitch. However, they failed to validate this subjective scale independently, so the result is questionable. Although subsequent validations of the model were carried out (Ekman, Goude and Waern, 1961, investigated darkness and subjective area; Eisler, 1960, investigated heaviness) they tested, in effect, a different model, as follows. Equation (2.1) can be rewritten as

$$g(x,y) = \frac{2 \frac{\min(x,y)}{\max(x,y)}}{1 + \frac{\min(x,y)}{\max(x,y)}} \quad (2.2)$$

so, writing

$$q_{xy} = \frac{\min(x,y)}{\max(x,y)},$$

we have

$$g(x,y) = \frac{2q_{xy}}{1+q_{xy}} \quad (2.3)$$

It was assumed in the latter two studies that subjective estimates of ratios of magnitudes would be equal to q_{xy} , so that if the relation (2.3) was found to be true, model (2.1) would be confirmed. Sjöberg points out that not only is this a questionable assumption, but also that Eisler in his 1960 study had to make several transformations of his ratio data to make the assumption hold.

There is thus good reason (see also Franzén, Nordmark and Sjöberg, 1972, for further evidence) to look at other possible models of unidimensional similarity. The most likely alternative seems

to be that g is closely related to q_{xy} - either by identity (Sjöberg, 1973; Hoijer, 1969a, 1970b; Waern, 1968a) or by a power law (Künnapas and Künnapas, 1971, 1973). Neither Sjöberg nor Waern tested the identity relation however, and Hoijer (1970b) in fact needed additional parameters to make it fit; on the other hand, Fisler and Ekman, in their original (1959) study, found evidence for rejecting it. The power law relation thus seems more reasonable:

$$g(x,y) = q_{xy}^{\alpha} \quad (2.4)$$

Künnapas and Künnapas present evidence supporting this and suggest in the 1973 paper that the ratio estimation is also a power function of q . This agrees with Luce's (1961; but see also Luce and Galanter, 1963b, pp. 255-6) choice theory analysis of similarity and Luce and Galanter's (1963b, pp. 288-289) mathematical analysis of the form of the generalisation function. Both of these analyses are based on questionable assumptions however: see sections 2.1.3.3 and 2.1.3.4 for further details.

We are therefore in the position that neither model is very convincing, but, equally, neither is very far wrong. Most multidimensional content models are designed to reduce to (2.1) in one dimension; it would seem prudent to develop a parallel series reducing to (2.4).

The unidimensional model was generalised by analogy: it was noted (Ekman, 1961; Sjöberg, 1973) that the numerator, $2\min(x,y)$ could be taken to represent the "sensory experience" common to the two stimuli, while the denominator, $x+y$, could represent the total "sensory experience". Thus similarity could be thought of as the ratio between the common, C_{xy} , and the total, T_{xy} :

$$g(x,y) = \frac{C_{xy}}{T_{xy}} \quad (2.5)$$

It remained to specify C_{xy} and T_{xy} . Two basic avenues were followed. The main one was the "vector" representation. Here a stimulus was thought of as a vector in \mathbb{R}^{n+} (see Figure 1.1), its intensity being represented by its vector length, h_x and "qualitative variation" between stimuli x and y being represented by the angle ϕ_{xy} between them. Their projections on each other

could be used in various ways to represent "common sensory experiences". Note that since all these constructs are defined in, and are largely meaningful only in, Euclidean space, the vector models represent a type of embedding in Euclidean space.

Some of the more well-known vector models are given in Table 2.1, along with their associated definitions of "common" and "total". Their variety and the apparently contradictory character of some of these definitions (compare, for example, that of Ekman et al., 1963, with the first of Ekehammar, 1972a) is symptomatic of their weak intuitive basis. It should be also noted that all reduce to (2.1) in the unidimensional case.

That of Goude (1966) is somewhat different in form to the others. Instead of defining similarity in the usual content way, Goude regarded it as complementary to a difference measure. He defined difference to be the ratio of what was not in common (taken to be the distance between vector termini) to the total.

Some of the models have had mathematical flaws, as pointed out by Sjöberg (op.cit).

TABLE 2.1
SOME VECTOR CONTENT MODELS OF SIMILARITY

Notation as in Table 1.1 and Figure 1.1 with

$$\begin{aligned} C_1 &= \min(p_{yx}, h_x) + \min(p_{xy}, h_y) & T_1 &= h_x + h_y \\ C_2 &= 2\min(p_{xy}, p_{yx}) & T_2 &= \sqrt{h_x^2 + h_y^2 + 2h_x h_y \cos \phi_{xy}} \\ C_3 &= T_2 \cos \phi_{xy} & S(x, y) &= C/T \end{aligned}$$

Definitions of:

Common Elements	Total Elements	$S(x, y)$	Qualitative Variation Only ($h_x = h_y$)	Reference
C_1	T_1	$\frac{\min(h_y, h_x \cos \phi_{xy}) + \min(h_x, h_y \cos \phi_{xy})}{h_x + h_y}$ ($= \cos \phi_{xy}$ if $p_{xy} \leq h_y$ and $p_{yx} \leq h_x$)	$\cos \phi_{xy}$	Ekman and Lindman (1961)
C_1	T_2	$\frac{\min(h_y, h_x \cos \phi_{xy}) + \min(h_x, h_y \cos \phi_{xy})}{\sqrt{h_x^2 + h_y^2 + 2h_x h_y \cos \phi_{xy}}}$	$\frac{\cos \phi_{xy}}{\cos(\phi_{xy}/2)}$	Ekman, Engen, Kunnepas and Lindman (1963)
see text	T_2	$1 - \frac{\sqrt{h_x^2 + h_y^2 - 2h_x h_y \cos \phi_{xy}}}{\sqrt{h_x^2 + h_y^2 + 2h_x h_y \cos \phi_{xy}}}$	$1 - \tan(\phi_{xy}/2)$	Goude (1966)
C_2	T_2	$\frac{2\min(h_x, h_y) \cos \phi_{xy}}{h_x + h_y} \quad (*)$	$\cos \phi_{xy}$	Eisler (1967)
C_3	T_1	$\frac{\cos \phi_{xy} \sqrt{h_x^2 + h_y^2 + 2h_x h_y \cos \phi_{xy}}}{h_x + h_y}$	$\cos \phi_{xy} \cos(\phi_{xy}/2)$	Ekehammar (1972a)
C_3	T_2	$\cos \phi_{xy}$	$\cos \phi_{xy}$	Ekehammar (1972a)

(*) Eisler also introduced a parameter m such that the factor $\cos \phi_{xy}$ was raised to the m th power. There seems little justification for this additional parameter.

For example, that of Ekman and Lindman (1961) is insensitive to changes in quantitative variation in parts of the space, while Ekehammar's second model (Ekehammar, 1972a) is totally insensitive to intensity. For vectors at a fixed angle, Ekehammar's first model gives a minimum similarity (intuitively, it should be a maximum) when the vector lengths are equal.

The other avenue followed in the interpretation of "common" and "total" is perhaps more intuitively obvious and straightforward, but, for practical reasons, less used. This considers stimuli primarily as sets of aspects or properties. Then, by defining a measure, \underline{m} , (for example, but not necessarily, the counting measure) on these aspects, we can define

$$C_{xy} = \underline{m}(X \cap Y).$$

The most logical definition of totality would be

$$T_{xy} = \underline{m}(X \cup Y) \quad (2.6)$$

and this has been recognised by Gregson (1970), although he uses it in a slightly different form.

However, Eisler, obviously generalising the unidimensional model of similarity, equation (2.1), takes

$$T_{xy} = \underline{m}(X) + \underline{m}(Y) \quad (2.7)$$

(see Eisler, 1964, 1967).

Thus we have the two similarity models respectively,

$$g(x,y) = \frac{\underline{m}(X \cap Y)}{\underline{m}(X \cup Y)} \quad (2.8)$$

and

$$g(x,y) = \frac{2\underline{m}(X \cap Y)}{\underline{m}(X) + \underline{m}(Y)} \quad (2.9)$$

The question is now, how to define \underline{m} so that equations (2.8) and (2.9) can be applied in practice. Eisler (1967) avoids the issue by merely decomposing the right hand side of equation (2.9) into "communalities",

$$\frac{\underline{m}(X \cap Y)}{\underline{m}(X)} \quad \text{and} \quad \frac{\underline{m}(X \cap Y)}{\underline{m}(Y)}$$

(cf. Bush and Mosteller, 1951; and section 2.1.3.5).

Gregson, in a treatment that does little to clarify the question, assumes a structure on the space that is equivalent to a dimensional structure, so that he can write

$$\mathcal{J}(x,y) = \frac{\sum_{i=1}^n \underline{m}(X_i \cap Y_i)}{\sum_{i=1}^n \underline{m}(X_i \cup Y_i)}$$

and assumes that \underline{m} is the counting measure, although it is difficult to see how this can be applied in continuous spaces. He then states that

$$\begin{aligned} \underline{m}(X_i \cap Y_i) &= \min[\underline{m}(X_i), \underline{m}(Y_i)] \\ \text{and } \underline{m}(X_i \cup Y_i) &= \max[\underline{m}(X_i), \underline{m}(Y_i)] \end{aligned}$$

(which is only true if $X_i \subseteq Y_i$ or $Y_i \subseteq X_i$; i.e. a prothetic continuum, or nesting sequence of sets), so that

$$\mathcal{J}(x,y) = \frac{\sum_{i=1}^n \min[\underline{m}(X_i), \underline{m}(Y_i)]}{\sum_{i=1}^n \max[\underline{m}(X_i), \underline{m}(Y_i)]} \quad (2.10)$$

Eisler (1964) has given a comparable treatment for the one dimensional case of his model. A clarification and extension of the arguments leading to this type of result will be the subject of Chapter 3.

Gregson also includes dimensional weighting in his model, but this could of course be incorporated in any dimensional model.

Roskam (1972) has also investigated the set theoretic approach using Eisler's model, and states definitions that would lead to a similar result to equation (2.10). He, too, fails to give any rigorous explanation for his definitions however. Another suggestion of Roskam's is that one should, instead of (2.9) put

$$\mathcal{J}(x,y) = \frac{\exp[\underline{m}(X \cap Y)]}{\exp[\frac{1}{2}(\underline{m}(X) + \underline{m}(Y))]}$$

where $\exp(a) = e^a$. This gives

$$\mathcal{J}(x,y) = \exp\left[-\frac{1}{2} \sum_{i=1}^n |\underline{m}(X_i) - \underline{m}(Y_i)|\right]$$

so that \mathcal{J} is a monotonic function of the City-Block distance. The attractiveness of this result seems

to be the only justification for its highly questionable derivation.

Three other models having a base in Content model theory are worth noting. Micko (1970) reinterpreted the vector content models using the locus of all projections of a vector onto a variable vector of constant length - the "halo" of the vector - as a measure of the stimulus it represents. He could define "common" and "total" in a way quite analogous to the set theoretic approach, and his final result, using Eisler's interpretation of the set-theoretic content model (equation 2.9) was

$$g(x,y) = 1 - \frac{d(x,y)}{d(x,e) + d(y,e)} \quad (2.11)$$

where d could be any distance. In the case of the Euclidean distance, this becomes

$$g(x,y) = 1 - \frac{\sqrt{h_x^2 + h_y^2 - 2h_x h_y \cos \phi_{xy}}}{h_x + h_y}$$

$$= 1 - \frac{\left(\sum_{i=1}^n (x_i - y_i)^2 \right)^{\frac{1}{2}}}{\left(\sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}} + \left(\sum_{i=1}^n y_i^2 \right)^{\frac{1}{2}}} \quad (2.12)$$

Hoijer (1969a,b) has put forward an additive model based on the assumption that the unidimensional similarity judgment is equal to the ratio judgment:

$$D(x,y) = 1 - \frac{1}{n} \sum_{i=1}^n \frac{\min(x_i, y_i)}{\max(x_i, y_i)} \quad (2.13)$$

Its objectionable properties include its apparent assumption that D equals $1 - \beta$, and its additive nature.

Waern (1968a,b, 1969, 1970b, 1971a,c) has proposed a multiplicative model of similarity

$$\beta(x,y) = \prod_{i=1}^n (\beta(x_i, y_i))^{n_i}$$

where $\beta(x_i, y_i)$ is the unidimensional similarity on the i th dimension. If

$$\beta(x_i, y_i) = \left(\frac{x_i}{y_i} \right)^{\alpha_i} \quad (x_i \leq y_i)$$

then $-\log \beta(x,y)$ is the (weighted) City-Block distance in log-transformed space. Waern has given it a Content model interpretation, and tested it,

but its major failing is that it is always zero when one of the points is on an axis.

2.1.1.2 Testing of the Models

While the original unidimensional similarity model, equation (2.1), was developed using a scale of the relevant continuum (although not independently validated), subsequent tests have used the equation (2.3), effectively only testing the relationship between ratio judgments and similarity judgments. This method was generalised for the multidimensional vector models, adhering to the principle that psychological phenomena could be usefully investigated only at the "proper level" (Ekman, 1961). However, this in itself involved yet another modelling assumption: that the multidimensional ratio judgment (viz., "report the proportion (q) of a standard percept that is contained in a given percept, and vice versa": Ekman, 1963) was equal to the ratio of the projection on a vector and the length of that vector:

$$q_{xy} = \frac{p_{xy}}{h_y} = \frac{h_x}{h_y} \cos \phi_{xy} \quad (2.14)$$

From such judgments, since

$$q_{xy}q_{yx} = \cos^2 \phi_{xy},$$

the angles ϕ_{xy} could be estimated, and since

$$\frac{q_{xy}}{q_{yx}} = \frac{h_x^2}{h_y^2}$$

the lengths h_x could be estimated (subject to certain normalisations - see, for example, Ekman, 1963; Bäckström and Goude, 1966; Eisler, 1967; Lund, 1974). Although these assumptions have considerable empirical support (e.g. Ekman, 1965), they can produce rather unorthodox results (e.g. Ekman, 1963), and put an extra, seemingly unnecessary, burden of modelling assumptions onto similarity modelling.

Not only have these ratio modelling assumptions often been incorporated into tests of the similarity models, but, even where this has not been so, the models have usually been tested in only the simpler, special case of purely "qualitative" variation: i.e. it is assumed (see fourth column of Table 2.1) that all vector lengths are the same, so that the only variation is in angle. Stimuli

are chosen, and/or subjects instructed, so that they will (hopefully) perform accordingly. Examples can be found in Ekehammar (1972b); Ekman (1965); Ekman et al. (1963, 1964). Some of their results do not appear to bear out the equal-length assumption (especially Ekman, 1965, and Ekehammar, 1972b). An additional difficulty is that, as can be seen from Table 2.1, most of the models are not very different under only "qualitative" variation, so this is not a good way to compare models.

However, an important underlying structure seems to be revealed by this simplification. To test the model with qualitative variation only, stimuli are usually chosen that, in an intuitive way, vary only qualitatively: Ekman et.al. (1963, 1964), and Ekman (1965) use words denoting emotional states or personality traits, smells of the same intensity, and colours of constant brightness; Ekehammar (1972b) uses Rorschach cards; while geometric figures and letters have also been used (e.g. Ekehammar, 1972a). These all represent continua which are often thought of as "metathetic" (qualitative, substitutive). Therefore, what is really happening is that it is being recognised that a different model is needed for metathetic continua than for quantitative (prothetic) continua. Thus

equations (2.1) or (2.4) apparently are not expected to apply to unidimensional metathetic continua. The main exception appears to be in the continuum of pitch, which is usually taken to be metathetic, and on which continuum equation (2.1) was originally tested. Evidence against (2.1) on the pitch continuum was found by Franzen et.al. (1972), so it remains controversial. Some discussion relevant to the applicability of (2.1) to different types of continua can be found in Eisler (1960, 1964), and Ekman, Goude and Waern (1961).

On the other hand, in practice unidimensional metathetic continua rarely occur in the results of the relevant experiments mentioned. They are always split up into two or more, apparently prothetic, continua. Either the method of data analysis (usually Factor Analysis) is at fault, or the models do not account for metathetic continua very suitably. (See also Ekman and Sjöberg, 1965; Torgerson, 1965; and Micko, 1970 for comments on this problem).

The surprisingly good fit of Ekman and Lindman's model can be interpreted in the light of this

discussion: in a large part of the space it reduces to

$$g(x,y) = \cos\phi_{xy}$$

and is often used in this form regardless. This is, or is nearly, just the form of the "qualitative" special case of the other vector models. Two experiments giving good agreement with the model - Ekman and Lindman (1961), and Lund (1974) - in fact only test the model in spaces that could easily be regarded as "qualitative". Thus Ekman and Lindman's model may not be valid for spaces of prothetic continua.

An additional difficulty with using ratio judgments is that they are rather difficult for the subject to understand and give. In fact Lund (1974) suggests that subjects cannot follow the instructions correctly and "tend to interpret all variations as qualitative". This may possibly be cleared up by use of the prothetic/metathetic distinction as above; or it may be the case that the findings of Künnapas and Künnapas (1971, 1973), that unidimensional similarity and ratio judgments are in fact almost the same thing, is also true for the multidimensional case.

Thus, the vector models have many problems. In his major study comparing most of the vector Content models so far suggested, Lund (1974) concluded that that of Ekman and Lindman was the best, but that they all had basic inadequacies, for both theoretical and practical reasons. Most findings are in accord with this: no vector model is very good; but equally none is very far from either the data or each other. The conclusions of Ekehammar (1972b, testing his two models and those of Ekman and Lindman, and of Ekman et.al., 1963) that the models "give only negligible differences in outcomes for the same data"; and of Waern (1970b, testing, among others, the models of Gregson, Ekman et.al. 1963, and Eisler, 1967) that "all similarity models here compared are equally far from the 'true' similarity model", are typical. The models capture enough of the data properties to be worthwhile, but not enough to be good models.

Gregson's model has rarely been tested, except in the so-called "Quadratic Similarities" situation (Gregson, 1970, 1972), and here the results were not very encouraging. However, in her comparative study, Waern (1970b) concluded that Gregson's model was one of the best of a rather mediocre collection.

An unpublished study on "similarity learning" (Field, 1974) and one by Rosenberg (1974) also give indirect support to the model. A general method for testing the model has been to substitute physical stimulus descriptions for the $\underline{m}(X_i)$, $\underline{m}(Y_i)$ terms (see equation (2.10)) and solve for dimensional weights. As Waern points out, this is a strong test of the model if it is successful. The use of magnitude estimations in this way, although it imposes a fixed dimensional structure on the stimuli, would seem to be the best way to test any such model with a minimum of modelling assumptions.

The other Content models mentioned here have also had only limited or no testing. Micko's has yet to be tested directly, probably due to the highly complicated instructions required, but a reanalysis of Ekman and Lindman's data using it gave fewer dimensions (Micko, 1970). Eisler's (1967) set theoretic model also had problems with instructions. Roskam's revision of Eisler's model has only been tested by use of MDS (Roskam, 1972), but he also compared the scaling results with magnitude estimation scale data in order to validate his findings.

Hoijer has tested his own model fairly extensively, including its additive structure, comparative dimensional weighting (Hoijer, 1969a) and order effects on similarity (Hoijer, 1970a). He also claims to have investigated its metric or non-metric nature (Hoijer, 1969b, 1970b), but this is to a large extent invalidated by his confusion between metrics and distances (see Hoijer, 1969b).

Waern's model is fairly easily tested, at least for its superficial properties, since it requires only unidimensional and multidimensional similarity estimations. This assumes that the dimensions are known to the experimenter and obvious to the subject, so that greater subtlety would be needed to test it more thoroughly. Support for her model is equivocal when she tests it with these (and other) considerations in mind: Waern (1968a,b, 1969, 1970b).

Finally, it should be noted that all the Content models, with the occasional exception of Gregson's and Hoijer's, have been tested only on grouped data.

2.1.2 DISTANCE^{*} MODELS

2.1.2.1 Models

The Euclidean Distance model was the first model used in quantitative similarity research. It underlay Thurstone's work, and Richardson (1938) used it specifically through the techniques of Factor Analysis made possible by the theorems of Young and Householder (1938) (see also Torgerson, 1958, Chapter 11; Gower, 1966). It was also the distance used by Coombs, although he recognised that others were theoretically possible (Coombs, 1964, p. 202).

The first worker to question its use was apparently Attneave, who compared it with the purely additive City-Block distance which had previously been given a physiological basis by Landahl (1945):

*Note that "Distance" is not used here in the strict sense of Chapter 1, but in the rather indefinite sense of Ekman and Sjöberg (1965).

$$D(x,y) = \sum_{i=1}^n |x_i - y_i|$$

He found (Attneave, 1950) that this distance fitted better than the Euclidean one, but since the study compared only the two models, it did not positively confirm either of them.

In the search for a more general Distance, the Minkowski (or ℓ_p) distance was suggested (e.g. Torgerson, 1958, p. 293; Kruskal, 1964a):

$$D(x,y) = \left(\sum_{i=1}^n |x_i - y_i|^\alpha \right)^{1/\alpha}, \quad \alpha \geq 1 \quad (2.15)$$

This is the City-Block distance when $\alpha = 1$ and the Euclidean distance when $\alpha = 2$.

Micko and Fischer (Micko and Fischer, 1970; Fischer and Micko, 1972) have given another generalisation of Distances by constructing them from a function on the directions in \mathbb{R}^n , interpretable as an "attention distribution" in the space. The Minkowski distances are special cases of these distances, which include most distances (the "(general) Minkowski metrics") in the sense defined in Chapter 1. Micko and Fischer (1970, p. 128)

point out that "it seems likely that subjective spaces are not always [general] Minkowskian" because of the assumptions it implies about their affine structure (D5 and D6 of Chapter 1). They discuss such distances, it appears, only because they are mathematically more tractable than a general metric; but even then they give little justification for assuming the triangle inequality (D7).

A different generalisation comes from suggestions (e.g. Shepard, 1960; Indow, 1974) that, even if psychological distance is not Euclidean, it may be at least locally Euclidean - i.e. Euclidean within small areas (volumes, hypervolumes) of the space. This suggests that subjective space may be a Riemann (Indow) or Finsler (Shepard, 1964) space. These are locally Euclidean; an example (see Shepard, 1960) is the surface of a sphere in Euclidean space. Except in the study of the metric of binocular visual space (e.g. Luneberg, 1950; Blank, 1953, 1958), these metrics do not appear to have been taken seriously enough to warrant testing.

The widest generalisation of Distances has been due to the axiomatisation by Beals, Krantz and

Tversky (Beals and Krantz, 1967; Beals, Krantz and Tversky, 1968; Tversky and Krantz, 1970) of distance functions. They noted that the Minkowski (power) metric has such properties as decomposability (distance is a function of independent contributions of components), intradimensional subtractivity (each such contribution is a function of the absolute value of the scale difference), interdimensional additivity (distance is a function of the sum of the contributions), and segmental additivity (there is a path between any two points on which the distances are additive), in addition to the defining properties of a metric. They found necessary and sufficient ordinal conditions for similarity or distance to be monotonic with such distance functions, and show that the Minkowski metric is "the only additive difference metric which is also a metric with additive segments" (Tversky and Krantz, op.cit., p. 588). Hence, these conditions are critical to the acceptance of a Minkowski metric; however, they go further by saying that the properties of intradimensional subtractivity and interdimensional additivity "may be taken as defining properties for psychological dimensions" because they are present in a "large class of multidimensional similarity models" (op.cit., p. 595). Since they have looked only at models that are distances, and mainly Minkowski distances, this conclusion seems rather rash. As we shall see, their

axiomatisation nonetheless provides a method that would not otherwise be available for testing these models.

Various other Distance models have been suggested but either seem to have little merit above those that they generalise or have not been tested in any generality (e.g. Gregson, 1966, 1968; Roskam, 1968, p. 29 and pp. 77-98; Young, 1973).

One other problem closely related to the use of Distance models should be noted here. Since Distance models model (if anything) dissimilarity, and similarity seems to be the more natural judgment for subjects to make (although dissimilarity judgments are of course quite often used - e.g. Hyman and Well, 1967, 1968; Tversky and Krantz, 1969), the function relating dissimilarity to similarity is a critical one if a model of similarity is to be constructed. We shall call this function the "**similarity gradient**" (cf. Shepard, 1958b). For Distance models, the similarity model consists of the distance function plus the similarity gradient. To a large extent the two functions are independent of one another. The main point of both the non-metric MDS programs, and the work of Beals, Krantz and Tversky, is of course to avoid this issue (although the former can

also throw light on it - see also Cunningham and Shepard, 1973), but for an accurate model of similarity it cannot finally be overlooked.

In so far as similarity is closely related to generalisation, the similarity gradient would seem to be closely related to the more classical generalisation gradient. Shepard (1957, 1958a,b) was one of the first to investigate this function rigorously. He found both empirical (but see Krantz, 1967) and convincing theoretical arguments supporting the exponential decay function:

$$g = e^{-D}$$

In particular, generalisation was an exponential decay function of Euclidean distance. Various other authors have followed him, supporting the exponential form of the similarity gradient, either pragmatically (e.g. Waern, 1968a,b, 1969, 1970a,b; Roskam, 1972) or for other, but often less convincing, reasons (e.g. Luce, 1961). Its main advantage seems to be that it is bounded between zero and one for positive arguments, so that the unbounded distance function can be mapped into the bounded similarity function (see Chapter 1). It also has pleasant and familiar mathematical properties, and both it

(e.g. Luce and Green, 1972) and its inverse, the logarithmic function, are probably the most commonly used non-linear functions in psychology. Helm, (see for example Coombs, 1964, p. 488; and Helm, Messick and Tucker, 1961) has also made the interesting observation that by exponentially transforming dissimilarity judgments, MDS results can be simplified, probably because Distance models overestimate large distances.

The other main contender is, predictably, the linear function. Most often it is used in the form (interpreted as "complement")

$$D = 1 - g$$

(e.g. Junge, 1960; Goude, 1966; Hoijer, 1969a,b; Eisler, 1967; Micko, 1970) which makes sense if D is also bounded, between zero and one, but is nonsense if, as is usual, D is a distance which is not scale-invariant; Waern (1970a,b) also points out that it may lead to violations of the triangle inequality. The alternative is to allow any linear transformation (with negative slope) - see for example Roskam (1972), Waern (1970b) -

$$D = -\alpha g + \beta \quad (\alpha, \beta > 0).$$

β must then be chosen so that D can at least attain its maximum value within the particular experiment. The theoretical objections to this are substantially no different to those for the "complement" transformation. Unless D is bounded, the linear gradient would not seem viable (see also Waern, 1970a,b; Rosenberg, 1974).

2.1.2.2 Interpretations and Testing

Since the Distance models have a mathematical origin, some effort has gone into discovering just what it means psychologically if one of them fits a certain body of data. The analyses of Micko and Fischer (1970) and Fischer and Micko (1972) show that most distances can be interpreted in terms of a particular attention distribution, by the observer, over the directions ("qualities") in multidimensional space. This has not been used in interpreting empirical results (and the analysis is not a unique one), but it largely coincides with findings on comparisons between the City-Block and Euclidean models. It has for some time been noticed (e.g. Torgerson, 1958, p. 292; Shepard, 1964) that the City-Block distance holds when the dimensions on which the stimuli vary are "simple, obvious and compelling" (Torgerson) or "analyzable" (Shepard). Euclidean distance was more suitable when the dimensions were less obvious or "homogeneous". The strongest support for this

is in two studies by Hyman and Well (1967, 1968). In a somewhat different, but not irreconcilable finding, Wender (1968) found that the Minkowski parameter increased with the difficulty of judging the similarity between stimuli.

The mere fact that an interpretation could be found for the parameter of the Minkowski models, however, is not necessarily a sign that the models are good ones: independent testing must be the final criterion. In fact, surprisingly little of this has been done. Until recently, studies of Distance models themselves - rather than studies in which different distance models are compared - were relatively infrequent. This is in contrast to research on Content models. On the other hand, researchers here have shown much more awareness of individual differences.

The best single study is probably that of Shepard (1964), who also reviewed some of the evidence for the Euclidean model. By using isosimilarity contours, with both direct similarity estimation and confusion probabilities, he was able to show that the triangle inequality was credible for individual subjects but not for group data. In another study using confusion probabilities, Krantz (1967) tested a "rational distance function" he had constructed from psychologically plausible postulates, and found

strong evidence to reject its being a metric (see section 2.1.3.3).

Many other studies tend to test the hypothesis that the model is Euclidean against the hypothesis that it is not. Although other tests, to identify any non-Euclidean model, are usually carried out (e.g. Attneave, 1950; Hyman and Well, 1967, 1968), these are more disconfirmations of the Euclidean model than positive evidence for any alternative. In addition it would often seem that the statistical methods used and/or the choice of the similarity gradient (see previous section), may confound the results. Luce and Galanter (1936b, p. 303) give a similar warning.

The two alternative methods are:

- (1) The use of non-metric MDS techniques; and
- (2) Investigation of underlying properties of the Distance models.

The first alternative is in many ways not much better than that previously discussed: it can consist merely of a comparison between the lack of comfort (stress) of two Procrustean beds. It can, however, be used in a more positive way, especially if it is used in comparison with non-Minkowskian models (Lund, 1974; Roskam, 1972; Eisler, 1967).

The technique, on its own, usually compares goodness-of-fit values, numbers of dimensions, and interpretability of dimensions. This may be very unreliable, as different models can give equally good fit on all these criteria - see section 2.2. Bearing this in mind, strong but equivocal support is given to the City-Block and Euclidean distances by this method.

The second alternative is largely motivated by the mathematical work of Beals, Krantz and Tversky. Some of the ordinal conditions they give for existence of a suitable distance can be tested experimentally, and they have given a lead in doing this. Tversky and Krantz (1969) tested interdimensional additivity and "strongly confirmed it" in the context of their experiment. Wender (1971) tested decomposability and intradimensional subtractivity and rejected them.* In the same spirit, Fenker (1972) gave a counter-example to interdimensional additivity. Again, support for the Distance models is equivocal. Becker and Pipahl (1974) have suggested this may be due to the perceptual structure of the stimuli, so that Distance models would apply only in certain contexts: where there is a "nested" structure.

* Similar results were found by Krantz and Tversky (1975).

Finally, a few content model workers have tested Minkowski models by combining unidimensional similarity judgments and by other cleverly contrived model testing procedures (e.g. Waern, 1968b, 1969, 1970b; Hoijer, 1969a). Hoijer (1969b, 1970b) has also tried to use isosimilarity contours to distinguish between various models, including Distance ones. His results were largely invalidated by his confusion between metrics and distances rendering incorrect much of his contour analysis.

In summary, Distance models, despite their widespread use, have yet to be fully tested. What rigorous tests have been done, are by no means unanimous in their support, and there are clear cases for rejection of the models. The similarity gradient still needs to be fully investigated.

2.1.3 OTHER MODELS

The models in this section are from other branches of psychology or other sciences. In some cases they do not pretend to be models of psychological similarity; in the others their validity relies largely on that of the underlying theory from which they arise.

2.1.3.1 A Set Theoretic Model

Restle, in his set theoretic treatment of judgment and choice behaviour (Restle, 1959, 1961) suggested that similarity was a measure of the aspects common to two stimuli:

$$s(x,y) = \underline{m}(X \cap Y)$$

(Restle, 1961, p. 32) for some \underline{m} such that: for all X, Y

$$(i) \quad \underline{m}(X) \geq 0;$$

$$(ii) \quad \underline{m}(\emptyset) = 0;$$

$$(iii) \quad X \cap Y = \emptyset \Rightarrow \underline{m}(X \cup Y) = \underline{m}(X) + \underline{m}(Y),$$

although (iii) is usually replaced with

$$(iv) \quad (\forall i \neq j, X_i \cap X_j = \emptyset) \Rightarrow \underline{m}\left(\bigcup_{i=1}^{\infty} X_i\right) = \sum_{i=1}^{\infty} \underline{m}(X_i).$$

(op.cit., p. 10). In addition, he assumed the distance between two stimuli should be a metric, and defining

$$D(x,y) = \underline{m}(X \Delta Y)$$

(Restle, 1959; op.cit., p. 43) he was able to show that D was a metric.

Restle also defined such constructs as linear arrays, betweenness, and dimensions (see Restle, 1959):

(a) Y is between X and Z (b_{xyz}) iff

$$\underline{X \cap \bar{Y} \cap Z} = \bar{X} \cap Y \cap \bar{Z} = \emptyset$$

(In this case, $D(x, z) = D(x, y) + D(y, z)$)

(b) A linear array: if

$$X^* = \langle X_1, X_2, \dots, X_n \rangle$$

$$Y^* = \langle Y_1, Y_2, \dots, Y_n \rangle$$

are two n -tuples of sets such that

$$(i) \quad X_i \subseteq X_j, \quad Y_i \subseteq Y_j \quad \forall i \leq j$$

(so that X^* , Y^* are families of nested sets);

$$(ii) \quad X_n \cap Y_n = \emptyset \quad \text{and} \quad X_n \cap Z = Y_n \cap Z = \emptyset,$$

then if $L_i = X_i \cup Y_{n-i+1} \cup Z$, $L^* = \langle L_i \rangle_{i=1}^n$ is

a linear array of sets. A linear array is the only array with b_{ijk} for all members with $i \leq j \leq k$. Z is its core; X^* , Y^* are its polar arrays. Two linear arrays are parallel if they differ only in their cores. A dimension is a set of parallel linear arrays. These may have cores that themselves form linear arrays; in which case we can have an n-space.

With this apparatus, it can be shown that D is in fact the City-Block distance (see Becker and Pipahl, 1974).

Thus Restle's models are much more closely allied to the Distance models than to the Content models; this is mainly because he emphasises the need for D to be a metric and for it to be additive on arrays; in fact he rejects a content-model-like distance function,

$$D(x,y) = \frac{\underline{m}(X\Delta Y)}{\underline{m}(X\cup Y)},$$

put forward by Galanter (1956), for not being additive.

This all side-steps the question of the validity of a set theoretic model of behaviour. Luce and Galanter (1963b, p.297) comment that "it seems no more intuitively acceptable to us to assume the existence of sets of aspects and a measure over them than to assume directly the existence of distances between pairs of stimuli, which is what the aspects and measure are intended to justify"; and again (p. 303), "Because we have no experimental identification either of aspects or aspect measures, it is anyone's guess whether [Restle's] notion of distance has any relation to those that have arisen in the response models". These views may be somewhat extreme: Gregson's similarity model and Tversky's (1972) or Restle's (1961) theories of choice seem to be indirect evidence for the validity of the overview set theory supplies. Certainly, we do not have to accept Restle's particular notion of distance. In the future, set theory could conceivably be given a more direct interpretation - for example, in terms of neural excitation.

In this section it could additionally be briefly noted that the models of Arabie and Boorman (Arabie and Boorman, 1973; Boorman and Arabie, 1972), while defined on sets of stimuli instead of sets of aspects,

and not strictly relevant to the present research, have some similarity to Restle's.

2.1.3.3 Models from Choice Theory

In his theory of individual choice behaviour, Luce (1959b, p. 47 ff.) considered interactions of continua, and in a subsequent paper (Luce, 1961) spelt out some of the consequences of his choice theory for similarity theory. Although his theory of choice has been found inadequate in many situations (e.g. Rumelhart and Greeno, 1971; Tversky, 1972), his analysis will still be valid in some circumstances and is one of the very few axiomatic derivations of a similarity model.

The theory is formulated in terms of $P_T(x;a)$, the probability that x will be chosen, from the set T , to be most similar to a . By Luce's choice axiom, we can write

$$P_S(x;a) = \frac{v(x,a)}{\sum_{y \in S} v(y,a)} \quad (2.16)$$

for any $S \subseteq T$. The scale v can behave like a

similarity scale, and we shall henceforth denote it by β . Conditions are found for symmetry in its arguments. Luce then asserts, giving little justification, that β must be related to a metric which is additive in one dimension. After rejecting $1/\beta$ as a possible transformation, he shows that $-\log\beta$ has reasonable properties. Next, by assuming that discrimination probabilities satisfy the choice axiom, Luce shows that, in one dimension,

$$\beta(x,y) = \frac{\min(v(x), v(y))}{\max(v(x), v(y))} \quad (2.17)$$

where the one-place v scale denotes the discrimination scale. Subsequently, Luce and Galanter (1963b, p.255) questioned the assumption leading to this result, and suggested without further comment that

$$\beta(x,y) = \left[\frac{\min(v(x), v(y))}{\max(v(x), v(y))} \right]^\beta \quad (2.18)$$

where β was to be estimated from the data, was more likely (cf. Künnapas and Künnapas, 1971, 1973; and section 2.1.1.1). Luce also gave a similarity analysis of bisection methods (see section 2.3).

Krantz (1967) generalised Luce's analysis by assuming that there exists a strictly increasing function F and a dissimilarity scale w , such that

$$P(x,y;z) = F[w(y,z) - w(x,z)] \quad (2.19)$$

where $P(x,y;z) = P_{\{x,y\}}(x;z)$. He made three other assumptions which were psychologically testable and which implied that the distance function

$$d(x,y) = \frac{1}{2}[w(x,y) + w(y,x) - w(x,x) - w(y,y)]$$

was a pseudometric. Taking

$$F(t) = (1 + e^{-t})^{-1}; \quad w = -\log v$$

we have Luce's choice model (equation (2.16)); it is also closely related to Shepard's (1957, 1958a) generalisation models (see section 2.1.3.5). With triadic data, Krantz was able to show conclusively that, at least for these choices of F and w , the two critical assumptions, including (2.19), implying that d was a pseudometric did not hold; neither did the triangle inequality hold. Krantz concluded that his attempt to construct a "rational distance function" (strictly, in our

terms, a "rational metric") had failed; he attributed it to its inability to "take into account the influence of 'irrelevant' dimensions". It does cast doubt onto both Luce's and Shepard's analyses.

Rumelhart and Greeno (1971) have also used similarity to explain certain faults in Luce's choice theory. Similarities were not defined however and were used merely as parameters to be solved for. Their final values showed great variation between subjects, which called the validity of their interpretation as similarity measures into question. The model tested has been largely superceded by Tversky's (1972).

A comment about the general validity of choice models of similarity is in order here. It is a question of fact whether or not the scales (e.g. of preference or discrimination) used in choice theory are closely enough related to, say, magnitude estimation scales, for choice theory models to be used in other contexts. A study by Cooper (1973) for example, indicated that preference space was somewhat different from similarity spaces. Other studies (Freedle, 1971; Steinheiser, 1970)

seem to indicate that the two spaces do not differ significantly in some situations.

2.1.3.4 Generalisation

The generalisation function plays an important part in theories of many areas in psychology and is obviously closely related to similarity. Luce and Galanter (1963b; pp. 284,288) give an analysis of the generalisation function as it relates to magnitude estimation scales in psychophysical scaling. They show (p. 288) the following:

If x, y, z are in one dimension and the generalisation function ζ is such that

1. $\zeta(x, y) = \zeta(1, y/x)x^\delta, \forall x, y > 0;$
 2. $\zeta(x, z) = \zeta(x, y)\zeta(y, z), \forall x, y, z \ni \text{either } x > y > z \text{ or } z > y > x;$
 3. $\zeta(x, y) = \zeta(y, x), \forall x, y > 0;$ and
 4. ζ is continuous in each of its arguments;
- then there exists a constant ζ such that

$$\zeta(x, y) = \left[\frac{\min(x, y)}{\max(x, y)} \right]^\delta$$

This obviously corresponds closely to their choice theory suggestion for one-dimensional similarity. Again, though, some of their assumptions seem (as yet) to have little psychological basis.

As mentioned before, Shepard (1957, 1958a,b) did a thorough empirical and theoretical investigation into the generalisation function.

Although some of his empirical work has been questioned (Krantz, 1967), his general findings were that generalisation is compatible with an exponential decay function of Euclidean distance.

Bush and Mosteller (1951), in a model for stimulus generalisation and discrimination, suggest the measure

$$g(x,y) = \frac{m(X \cap Y)}{m(Y)}$$

for similarity. This is asymmetrical (cf. Eisler's, 1967, "communality ratios" - see section 2.1.1.1) and thus does not seem a viable model of similarity.

A paper relevant to the cross-validity of discrimination models is Bechtel (1966).

2.1.3.5 Identification

In a choice model of identification (summarised in Luce, 1963; pp. 113-115), referring mainly

to work by E.F. Shipley in which similarity takes a central part, it is assumed that $-\log s$ is a metric - or that it behaves like a "psychological distance". Thus:

1. $\forall x, y, \quad s(x, y) = s(y, x)$
2. $\forall x, \quad s(x, x) = 1$
3. $\forall x, y, z, \quad s(x, z) \geq s(x, y)s(y, z)$

It is sometimes also assumed that if the stimuli can be viewed as having distinct components, then the overall distance is related to the distances on the components in the same way as a Euclidean distance. Luce comments that these assumptions "all arise from preconceived notions about the intuitive meaning of the s scale and from considerations of mathematical simplicity; they are neither obviously necessary nor clearly dictated by data, even though their consequences have received some empirical support". The same can be said of most of the constructions of section 2.1.3; an additional danger is that the "intuitions" and "preconceived notions" may all stem from "common-sense" experience of Euclidean distance so that the basic assumptions effectively become little more than a hidden assumption of Euclidean distance (cf. Luce and Galanter, 1963b, p. 303).

Luce also notes that the models in this area are closely related to Shepard's generalisation models.

2.1.3.6 Coefficients of Similarity

A huge variety of more or less arbitrary coefficients expressing similarity in various forms have been constructed over the decades. Most of them were not in any way designed to be formal models of psychological similarity, but they are of interest firstly because they tend to give some intuition into what various people have thought of as reasonable for expressing similarity (some are remarkably like various formal models of similarity), and secondly because many of them, when they are inner products, can be factor analysed. Hence they could conceivably serve, as have done the standard factor analytic models,

$$\sum_{i=1}^n x_i y_i$$

or

$$\sum_{i=1}^n x_i y_i$$

$$\frac{\sum_{i=1}^n x_i^2}{\sqrt{\sum_{i=1}^n x_i^2}} \frac{\sum_{i=1}^n y_i^2}{\sqrt{\sum_{i=1}^n y_i^2}}$$

as models of similarity (see Ekman and Lindman, 1961; but also Roskam, 1968, pp. 99-110). Some have been developed for clustering techniques (e.g. Gleser, 1968; Lance and Williams, 1966, 1967), others for taxonomy (e.g. Jaccard, 1912; Bray and Curtis, 1957; Goodall, 1966, Gower, 1967) and others are types of correlation coefficient (e.g. Cohen, 1969; Block 1970; Vegelius, 1973).

Vegelius (1973) gives an excellent survey of several correlation coefficients and their properties; but see also Cohen, Block, and Gleser.

More relevant to the present work are Bray and Curtis (1957) who used the coefficient

$$C = \frac{2 \sum_{i=1}^n \min(x_i, y_i)}{\sum_{i=1}^n x_i + \sum_{i=1}^n y_i}$$

where the $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$ are test scores on two different groups. Compare this with equation (2.9) and the discussion following it.

Lance and Williams discuss the Bray-Curtis coefficient along with some more conventional ones in their 1966 paper, and in their 1967 paper discuss also the "Canberra metric"

$$\sum_{i=1}^n \frac{|x_i - y_i|}{x_i + y_i} \quad \text{or} \quad \sum_{i=1}^n \frac{|x_i - y_i|}{|x_i| + |y_i|}$$

(cf. Gregson, 1966; Hoijer, 1969), and Gower's distance measure (Gower, 1967)

$$\sum_{i=1}^n \frac{|x_i - y_i|}{w_i}$$

where w_i is the range of attribute i .

A number of authors also discuss, in various forms, the measure (cf. equation (2.8))

$$S_{xy} = \frac{\underline{m}(X \cap Y)}{\underline{m}(X \cup Y)}$$

or its complement

$$1 - S_{xy} = \frac{\underline{m}(X \Delta Y)}{\underline{m}(X \cup Y)}$$

(Galanter, 1956; Gibson, 1965; Jaccard, 1912).

A notable feature common to all these coefficients is their normalised nature; this differentiates them sharply from the distance model type of measure. The more sophisticated coefficients (see especially Goodall, 1966; but also Gower, 1967, Lance and Williams, 1967) also recognise problems that are shared with models of similarity: how to decide which dimensions to include in the measure; the related problem of how to weight the dimensions, especially when they are over different ranges or have different units; and how to treat discrete, non-quantifiable data.

2.1.4 COMPARISONS AND CONCLUSIONS

One of the few investigations not suffering from many of the faults of testing outlined in sections 2.1.1.2 and 2.1.2.2 is that of Waern (1970b) who compared the following models: City Block and Euclidean (with both exponential and linear transformations), Ekman's et.al. (1963), Eisler's (1967), and Gregson's. By experimental design and various tests, she was able to conclude from five experiments involving two or three-dimensional stimuli, that City-Block distance was best, with Gregson's model a close second. The transformation could not be definitely decided.

Sjöberg (1973) in his largely theoretical comparison of the Distance and Content models, discusses usefully many of the problems in similarity unexplainable by either: the effect of different contexts and ranges of stimuli; the possibility of non-dimensional structures; the effect of instructions on the structure and the weighting of any dimensions; the relation of similarity to dissimilarity, oppositeness, identity, matching, and classification. He differentiates between the two classes of model on several grounds, including the problems of: scale consistency; interstimulus comparability; and dimensions unused by some subjects but used by others ("unique dimensions"). Sjöberg comes down in favour of the Distance models; his reasons seem to be largely ones of practicability however, true only because of the lack of clarity in the use of the Content models, and the great amount of mathematical development behind the Distance models. He does point the way, though, to some of the problems that will have to be solved to find a good model of similarity.

The most important of these problems are probably the effects of different contexts and ranges of stimuli, and the problem of unique

dimensions. The context problem is already being studied; the work of Hyman and Well shows that different models will be necessary in different contexts; it is yet to be seen how different they will eventually have to be. Ekman et.al. (1963, 1964); and Becker and Pipahl (1974) also discuss this. However, the considerations of these authors by no means cover all the problems of contextual effects. Torgerson (1965, p. 383) gave an example where doubling the stimulus range left the range of similarity judgments constant: similarity apparently changed according to the context. In certain circumstances this is easily explainable for the Content models by their property S6 (see Chapter 1); it is quite inexplicable according to the Distance models. There are still plenty of context effects the Content models do not explain (see Waern, 1971b; Wallach, 1958; Goldmeier, 1972; Ternus, 1938); far more sophisticated models than any considered here are obviously still needed.

However, the Content models do allow for an additional type of context effect not allowed for by Distance models - in the problem of unique dimensions. If "irrelevant" dimensions (usually

ones that do not vary over that sample) are present then the intradimensional subtractivity of the Distance models means that they cancel and may as well not be there. This does not happen with the Content models however; it may be that correct recognition of "irrelevant" dimensions may explain some problems of Context (see also Hoijer, 1959a; Goodall, 1966; Fenker and Brown, 1969; Shepard, 1964 for further references on "unique dimensions"). It may also partly explain why multidimensional scalings with Content models often have higher dimensionality than with Distance models (see section 2.1.1.2).

The other problems listed by Sjöberg have been studied in certain aspects by other authors not previously noted: for example, Johnson (1967), Waern (1971b), Cunningham and Shepard (1973), Arabie and Shepard (1973) on non-dimensional structures; Hoijer (1969), Fillenbaum (1973) on effect of instructions; Derks (1972), on matching; Bechtel (1966), Cooper (1973) on scale consistency; Fenker and Brown (1969), and Lockhead (1970) discuss considerations relating to stimulus comparability.

In summary, surprisingly little rigorous, detailed testing of models appears to have been done. On the Content model side, the only ones we can definitely count out are those of Ekman and Lindman (1961), and Ekehammar (1972a), for mathematical reasons. Paradoxically, on empirical grounds, that of Ekman and Lindman seems to be one of the best, though this may be only for spaces of metathetic continua. Gregson's appears promising, but not fully tested. The other Content models are only mediocre although they usefully reflect some properties of similarity. Thus the line of reasoning represented by Gregson's model seems worth developing.

The most likely on the side of the Distance models are the Minkowski distances with parameter from 1 to 2, inclusive, although few others have been seriously considered. Their main attraction has been mathematical properties such as additivity on straight lines, the triangle inequality, and the behaviour under affine transformations of the space. None of these properties as yet have much psychological support, though they are to a large extent untested. Their attractiveness may be due to one's Euclidean-based "intuitions" about what a distance should be like, and is therefore unreliable.

Problems also exist in that the two different schools tend to use different methods for testing their models, which to some extent make them difficult to compare. Ekman, for example emphasises testing models "at the proper level" (Ekman, 1961) and this has been largely adopted by the Content model school. Distance model testing tends to be only comparative and Procrustean, although the ordinal techniques of Beals, Krantz and Tversky are an exception. Different models may be necessary for prothetic and metathetic continua.

We might finally remark that none of the similarity models, except Hoijer's (1969a) and possibly Waern's (1968a) obey Luce's Principles of Theory Construction (Luce, 1959a). This is only typical of the generally chaotic state of the psychology of similarity.

2.2 PROBLEMS OF MULTIDIMENSIONAL SCALING

Since the beginnings of general use of MDS techniques (e.g. Torgerson, 1958), and particularly since the advent of the Shepard-Kruskal non-metric MDS algorithms, there has been a widespread and

growing awareness of their weaknesses, artefactual results, and technical problems. Shepard, Romney and Nerlove (1972) set out a wide range of techniques with many of their technical details and problems; we shall not be going into this here. The technical problems include the effects on the final configuration and its fit to the data of choice of initial configuration, stress formula, monotonic transformation and minimisation method. Such papers as Young (1973), Guttman (1968), Arabie (1973), Roskam (1970a,b) give some ideas on these. We are more concerned here with a question that assumes that these problems have been satisfactorily resolved, or at least that they can be made as irrelevant as possible by, for example, keeping to a fixed, reasonably satisfactory, set of such options. We shall be concerned with the effect of an algorithm's modelling assumptions: how do they affect the final configuration?

Generally, one would expect the model used to affect the solution in three ways:

- (1) It will affect the goodness of fit (stress);
- (2) It will affect the number of dimensions required for a reasonable fit;

- (3) It will affect the interpretability of the final configuration (and its dimensions).

It is generally assumed, at least in practice, that good results on all three of these criteria - namely, low stress, low number of dimensions, and good interpretability - is evidence for the validity of the model used. There are however, indications that this is not necessarily so.

Koopman and Cooper (1974) discuss an equivalence between the Minkowski α -distances with parameter, α , between 1.0 and 2.0, and those with parameter, α , between 2.0 and ∞ . They show that, in two dimensions, identical (in the case of $\alpha = 1$) or almost identical sets of interpoint distances can be reproduced from two different configurations, one in Minkowski α -distance space, the other in Minkowski α^* -distance space where

$$\alpha^* = \frac{\alpha}{\alpha-1}$$

Thus, for example, the City-Block ($\alpha = 1$) and "Dominance" ($\alpha^* = \infty$) spaces are in this sense equivalent, while Euclidean ($\alpha = \alpha^* = 2$) space has no equivalent Minkowski space. They also

show that the City-Block space in m dimensions has a corresponding Dominance space in 2^{m-1} dimensions with the property set out above.

Setting aside the question of interpretability of the dimensions or configurations of the alternative spaces (although one might cynically suggest that evidence seems to show that an interpretation can be found for almost anything - see Armstrong 1967), the implications of this are twofold. Firstly, a low stress for any particular Minkowski model in two dimensions does not mean necessarily that that model truly represents the way the subject is behaving, nor that the configuration it produces is the correct one: at least one alternative model and configuration exists. Secondly, a low dimensional solution for a City-Block model does not mean necessarily that the City-Block model is true or that the configuration is the correct one: the Dominance model with a higher dimensioned configuration is just as possible.

These implications not only have obvious practical importance, but they also serve as counterexamples to any supposed one-to-one relationship between models and configurations.

Good stress and low dimensionality are not sufficient evidence for the validity of a particular model. Attempts to determine "true" dimensionality or "true" goodness of fit by use of regressions on Monte Carlo data to find the relation between stress, number of points, and dimensionality (e.g. Young, 1968; Spence and Graef, 1974), until this point is resolved are merely begging the question.

Eisler (1967) and Roskam (1972) give a counter-example on the interpretability criterion. Two scalings of similarity data from an experiment by Eisler on simple point stimuli gave two different but equally interpretable two-dimensional configurations. (The stress values are not directly comparable since different stress formulas were used in each case). In other words, two different models gave quite different but equally plausible solutions. Thus not even interpretability in MDS solutions may be relied on as a guide to whether a model is correct.

We can conclude then, that on modelling assumptions alone - leaving aside any consideration of the better studied effects of technical options - none of the three usual criteria for acceptance of a model in MDS are reliable. Independent validation of the model is still needed.

2.3 PROBLEMS IN PSYCHOPHYSICS

In Psychophysics, as in MDS, there are large problem areas with which we shall not concern ourselves. Those of relevance here are Weber's Law and its proposed subjective counterpart, Ekman's Law; the prothetic versus metathetic scale distinction; and the relation between confusion, category and magnitude scales. Members of the content model school have made several attempts to relate the unidimensional similarity model to these problems, and we shall be re-examining some of them from this viewpoint in Chapter 5.

2.3.1 EKMAN'S LAW

Ekman (1961) suggested a subjective analogue of Weber's Law:

$$\Delta x = \alpha x + \beta$$

where β is near zero, which Stevens (1966) called "Ekman's Law". This has been given a similarity derivation (e.g. Ekman, Goude, and Waern, 1961) by assuming that "a j.n.d. is constant on the similarity continuum", so, using equation (2.1)

$$\alpha' = \beta(x, x+\Delta x) = \frac{x}{2x+\Delta x} \quad (2.20)$$

(α' a constant). Hence,

$$\Delta x = \frac{1-2\alpha'}{\alpha'} x$$

so, if

$$\alpha = \frac{1-2\alpha'}{\alpha'}$$

we have

$$\Delta x = \alpha x.$$

β is presumably to be fitted from data to account for the usual distortions for x near zero. A parallel argument holds if we assume that equation (2.4) holds instead of (2.1). The basic assumption is of course summarised in the lefthand equality of (2.20); it is open to experimental test, but it does not seem to explain much more than does Ekman's Law itself.

We assumed of course that x here was in subjective space - this would seem reasonable. If it is in physical space, the above "derivation"

gives Weber's Law. Also x must be on a prothetic continuum: see the comments in the following section.

2.3.2 PROTHETICNESS AND METATHETICNESS

Stevens and Galanter (Stevens and Galanter, 1957; Stevens, 1957) introduced a distinction between metathetic and prothetic continua, defining them largely in operational terms: the j.n.d. size increased up the scale on prothetic but not metathetic continua; category scales are concave downwards on magnitude scales on prothetic continua - they may be linear on metathetic continua; "time order error" exists only on prothetic scales; and the hysteresis effect probably only occurs on prothetic scales.

Restle (1959, 1961) has given a theoretical foundation to the two types of scale: he showed that the only types of scale with the "betweenness property" (see section 2.1.3.1) present everywhere were the substitutive or metathetic scales. The prothetic, quantitative, or intensive scale is a special case of the substitutive scale, in which nothing is substituted for what is removed: it forms a sequence of nested sets. Thus there is a continuum between metatheticness and protheticness (see also Eisler, 1963, 1964).

It is difficult to relate this interpretation directly to scaling. Gregson's similarity model is one attempt; another is found in Becker and Pipahl (1974), who derive the City-Block model from Restle's theory, assuming metathetic continua. They then try to interpret the stimuli in two experiments (Wender, 1971; and Tversky and Krantz, 1969) in view of the fact that only the second experiment found that psychological distance was consistent with a City-Block distance. They conclude that "the nested structure assumption" may be necessary for the decomposability condition to hold. While their interpretation is worthwhile, it is hardly unambiguous and needs further clarification.

2.3.3 CATEGORY VS MAGNITUDE SCALES

The operational criterion of protheticness that has been given the greatest attention by similarity theorists is that of the concave downwards relation between category scales and magnitude scales. In fact the relationship between the two scales varies a great deal, depending on various conditions including stimulus distribution or spacing, and stimulus discriminability down

the scale. For a logarithmic spacing of the stimuli, the relationship itself is almost logarithmic, as is a "pure" category scale suggested by Stevens and Galanter (op.cit.). The relationship between the magnitude and confusion scales is also logarithmic (e.g. Stevens, 1960b, 1966). Torgerson (1960) suggested a near-logarithmic relationship for category scales and rejected the discriminability condition; he also noted a complementary relationship between category scales of opposite dimensions and a reciprocal one between the corresponding pair of magnitude scales. From this observation and an additional assumption, Eisler (1962) showed that, if this was a general law, the logarithmic relationship must hold; he also gave an alternative derivation in terms of discriminability.

The mathematical purity of the logarithm also attracted derivations in terms of similarity judgments. The theory here says that categories are used by the subject to divide the stimulus range into intervals whose exemplars are evenly spaced on the similarity continuum. Thus if the range is $[x, y]$ and is to be divided into n categories, exemplified by x_1, x_2, \dots, x_n respectively, where $x = x_1$ and $y = x_n$, we have

$$\beta(x_i, x_{i+1}) = \beta(x_j, x_{j+1}) \quad \forall i, j \in \{1, 2, \dots, n\}.$$

Either of the unidimensional similarity models (equations (2.1) and (2.4)) then implies

$$\frac{x_i}{x_{i+1}} = \frac{x_j}{x_{j+1}} \quad \forall i, j \in \{1, 2, \dots, n\} \quad (2.21)$$

Ekman, Goude and Waern (1961) stated without proof that this result showed that the logarithmic relationship held; Junge (1960), with the help of the assumption of Steven's Power Law and the questionable assumption (which he calls a "banal truth") that $D = 1 - \beta$, gives a proof.

However, a related assumption to equation (2.21) gives questionable results. If we assume that the bisection judgment is also simply dividing the interval so that the bisection point is equisimilar to the two end points, then we have equation (2.21) for $n = 3$, so

$$\frac{x}{x_2} = \frac{x_2}{y}$$

where x_2 is the bisection point; so

$$x_2 = \sqrt{xy}.$$

Luce, (1961) points out that, even with hysteresis effects ignored, this value is too low, and overcomes this with bias parameters. Restle (1961, p. 216) in a related discussion gives evidence from Stevens and Galanter (op.cit.) that the middle category of a category scale is generally quite close to one-third of the (physical) distance from the smallest to the largest stimulus. He gives the rather unintuitive explanation that each stimulus consists of both "smallness" and "largeness" aspects; this, using the set theoretic interpretation, gives the required result. Restle's observation and the logarithmic relation are incompatible; for further discussion see Chapter 5.

One important point should finally be made: if the similarity analyses given in this section and in section 2.3.1 are at all correct, then since their results are true only for prothetic continua, the unidimensional similarity models on which the analyses are based can be true only for prothetic continua. A different model of (unidimensional) similarity will be needed for metathetic continua.

CHAPTER 3

A PSYCHOLOGICAL DISTANCE FUNCTION

The **distance models** are weak as to their empirical background; the **vector content models** are weak as to their mathematical tractability. In addition, the latter have such a weak intuitive structure that their resulting variety makes modelling a sophisticated curve-fitting exercise rather than the construction of a soundly based theory.

In this chapter we will construct a non-Minkowskian distance function with some basis in both set theory and empirical data. It will have an obvious but complex relationship to the Minkowski family. We will develop variations on existing models of similarity, and will show that the Minkowski distances do have a place in similarity modelling.

Our starting point will be the set-theoretic interpretation of the content models, for which Gregson's model has given some empirical foundation. To develop them, we shall follow Restle's use of set theory (see Restle, 1959, 1961), but extend it to an infinite case.

Sections 2.1.3.1 and 2.3.2 give a brief summary of some of his results and some warnings about their validity. We shall base the development on Gregson's interpretation of the content model (e.g. Gregson, 1970; and section 2.1.1.1) which will be denoted by \mathcal{G}_G :

$$\mathcal{G}_G(x,y) = \frac{\underline{m}(X \cap Y)}{\underline{m}(X \cup Y)} \quad (3.1)$$

and Eisler's (Eisler, 1967; and section 2.1.1.1) interpretation, to be denoted by \mathcal{G}_E :

$$\mathcal{G}_E(x,y) = \frac{2\underline{m}(X \cap Y)}{\underline{m}(X) + \underline{m}(Y)} \quad (3.2)$$

They will be developed in parallel in the context of the two different types of continua: prothetic and metathetic. Although Restle's analysis makes the first a special case of the second, we will find that the extremes of protheticness and metatheticness can give quite different models.

In all cases, we will be working in the space of "relevant aspects". This means in practice that aspects of the stimulus sample that are for some reason ignored by the subject will not be considered in our analysis.

An important property of the content models is that they are sensitive to all attributes present in the stimulus descriptions whether or not the stimuli all take the same value on a given attribute; thus an important question is how to decide which attributes are "relevant", or have an effect on the final judgment represented by the model. We are effectively begging this question by our choice of stimulus space. Further comments on this will be found in section 3.3.

In a given context, the measure m will be assumed to conform to the usual properties of a measure (see, for example Restle, 1959; or Halmos, 1950, p.30): properties (i), (ii), and (iv), as set out in section 2.1.3.1. It will represent a subjective assessment of the "size" of the sets of aspects of the stimuli. There may be inter-individual or temporal (e.g. Eisler, 1960, p. 78) variation in the measure, but we will not discuss this in detail. Effectively, our assumptions here represent a transfer from assumptions about the nature of the distance function to (perhaps more basic) assumptions about the nature of the measure on sets of aspects. In themselves they have little justification apart from the results they imply and their attractiveness mathematically and, perhaps, intuitively.

3.1 THE MODEL ON PROTHETIC CONTINUA

3.1.1 The Model

Consider a space Ω of "relevant aspects", constructed as follows:

Let there be n sequences of sets, Ω_i , $i=1, 2, \dots, n$, where

$$\Omega_i = \{A_{ia} : a \in \Omega_i\}$$

and $P.(a): \Omega_i \subseteq \mathbb{R}^+$

$$P.(b): a < a' \Rightarrow \text{either } A_{ia} \subseteq A_{ia'} \quad \forall a, a' \in \Omega_i,$$

$$\text{or } A_{ia'} \subseteq A_{ia} \quad \forall a, a' \in \Omega_i.$$

(i.e. Ω_i is a family of nested sets; in the first case it is an increasing sequence, in the second it is decreasing.)

$$P.(c): A_{ia} \cap A_{ja'} = \emptyset \quad \forall i \neq j; i, j \in \{1, 2, \dots, n\}, \\ \forall a \in \Omega_i, a' \in \Omega_j.$$

Then define the space as

$$\Omega = \{X: X = \bigcup_{i=1}^n X_i; X_i \in \Omega_i, i=1, \dots, n\}$$

This space is thus a set of unions of n disjoint prothetic continua. Each set in it is a unique union of sets, one from each continuum:

Suppose some set $X \in \Omega$ has two different such compositions - i.e. suppose one union is not unique; then

$$X = \bigcup_{i=1}^n X_i = \bigcup_{i=1}^n X'_i, \text{ where } X_i, X'_i \in \Omega_i$$

for $i=1, 2, \dots, n$

and $\exists j \in \{1, 2, \dots, n\} \ni X_j \neq X'_j$.

Take any such j . Then

$$\left(\bigcup_{i=1}^n X_i \right) \cap X_j = \left(\bigcup_{i=1}^n X'_i \right) \cap X_j;$$

i.e. $\bigcup_{i=1}^n (X_i \cap X_j) = \bigcup_{i=1}^n (X'_i \cap X_j)$ (3.3)

But, by $P_0(c)$, $X_i \cap X_j = X'_i \cap X_j = \emptyset$ if $i \neq j$.

Therefore (3.3) reduces to

$$X_j \cap X_j = X'_j \cap X_j$$

or $X_j = X'_j \cap X_j$, so $X_j \subseteq X'_j$.

By an exactly parallel argument with X'_j , we can show that

$$X'_j = X_j \cap X'_j, \text{ so } X'_j \subseteq X_j.$$

Thus $X'_j = X_j$ - a contradiction.

Hence $X'_i = X_i$ $i = 1, 2, \dots, n$

so the "two" decompositions are the same.

The space closely resembles a cartesian product of values in n dimensions. Assumption P.(a) above implies that each "dimension" can have either a finite or infinite number of members: that is, the dimension can be anything from discrete (discontinuous) and finite, to continuous. Assumptions P.(b) and P.(c) will be discussed later.

Taking any X in the space, then

$$X = \bigcup_{i=1}^n A_{i\alpha_i}$$

for some $\alpha_1, \alpha_2, \dots, \alpha_n$. However, we shall use the simpler notation

$$X_i = A_{i\alpha_i} \quad i = 1, 2, \dots, n.$$

so that

$$X = \bigcup_{i=1}^n X_i \quad \text{and} \quad X_i \cap X_j = \emptyset, \quad i, j = 1, 2, \dots, n, \quad i \neq j.$$

Hence, if \underline{m} is a measure as defined previously, then

$$\underline{m}(X) = \sum_{i=1}^n \underline{m}(X_i).$$

Now, since each \mathcal{C}_i is a family of nested sets, it is easily seen that

$$\underline{m}(X_i \cap Y_i) = \min(\underline{m}(X_i), \underline{m}(Y_i))$$

$$\text{and} \quad \underline{m}(X_i \cup Y_i) = \max(\underline{m}(X_i), \underline{m}(Y_i)),$$

for $i=1, 2, \dots, n$ and any $X, Y \in \mathcal{C}_i$.

But

$$\begin{aligned} X \cap Y &= \left(\bigcup_{i=1}^n X_i \right) \cap \left(\bigcup_{i=1}^n Y_i \right) \\ &= \bigcup_{i=1}^n (X_i \cap Y_i), \quad \text{using P. (c),} \end{aligned}$$

and

$$\begin{aligned} X \cup Y &= \left(\bigcup_{i=1}^n X_i \right) \cup \left(\bigcup_{i=1}^n Y_i \right) \\ &= \bigcup_{i=1}^n (X_i \cup Y_i) \end{aligned}$$

which are still disjoint unions. Hence

$$\underline{m}(X \wedge Y) = \sum_{i=1}^n \min(\underline{m}(X_i), \underline{m}(Y_i)) \quad (3.4)$$

and

$$\underline{m}(X \cup Y) = \sum_{i=1}^n \max(\underline{m}(X_i), \underline{m}(Y_i)) \quad (3.5)$$

Applying these to the defining equations (3.1) and (3.2), we get, respectively,

$$S_G(x, y) = \frac{\sum_{i=1}^n \min(\underline{m}(X_i), \underline{m}(Y_i))}{\sum_{i=1}^n \max(\underline{m}(X_i), \underline{m}(Y_i))} \quad (3.6)$$

$$S_E(x, y) = \frac{2 \sum_{i=1}^n \min(\underline{m}(X_i), \underline{m}(Y_i))}{\sum_{i=1}^n (\underline{m}(X_i) + \underline{m}(Y_i))} \quad (3.7)$$

We now simplify the notation by taking literally the cartesian product analogy mentioned above. Thus we put

$$x_i = \underline{m}(X_i), \quad i = 1, 2, \dots, n \quad (3.8)$$

and so on for all members of Ω . Note that $x_i \geq 0$ for all X and i . Rewriting (3.6) and (3.7) using this notation, we have

$$S_G(x, y) = \frac{\sum_{i=1}^n \min(x_i, y_i)}{\sum_{i=1}^n \max(x_i, y_i)} \quad (3.9)$$

and

$$\beta_E(x, y) = \frac{\sum_{i=1}^n \min(x_i, y_i)}{\sum_{i=1}^n (x_i + y_i)} \quad (3.10)$$

Now it is easily seen that for any $\alpha, \beta \in \mathbb{R}$,

$$\min(\alpha, \beta) = \frac{\alpha + \beta - |\alpha - \beta|}{2} \quad (3.11)$$

$$\max(\alpha, \beta) = \frac{\alpha + \beta + |\alpha - \beta|}{2}, \quad (3.12)$$

so substituting back into (3.9) and (3.10),

since $x_i, y_i \in \mathbb{R} \quad \forall i \in \{1, 2, \dots, n\}$, we have

$$\beta_G(x, y) = \frac{\sum_{i=1}^n [x_i + y_i - |x_i - y_i|]}{\sum_{i=1}^n [x_i + y_i + |x_i - y_i|]}$$

and

$$\beta_E(x, y) = \frac{\sum_{i=1}^n [x_i + y_i - |x_i - y_i|]}{\sum_{i=1}^n (x_i + y_i)}$$

If we divide both numerator and denominator of both these equations by $\sum_{i=1}^n (x_i + y_i)$, we find:

$$\beta_G(x,y) = \frac{1 - \frac{\sum_{i=1}^n |x_i - y_i|}{\sum_{i=1}^n (x_i + y_i)}}{1 + \frac{\sum_{i=1}^n |x_i - y_i|}{\sum_{i=1}^n (x_i + y_i)}} \quad (3.13)$$

and

$$\beta_E(x,y) = 1 - \frac{\sum_{i=1}^n |x_i - y_i|}{\sum_{i=1}^n (x_i + y_i)} \quad (3.14)$$

(Note that if $\sum_{i=1}^n (x_i + y_i) = 0$, both models are undefined).

It is now obvious that the function

$$D(x,y) = \frac{\sum_{i=1}^n |x_i - y_i|}{\sum_{i=1}^n (x_i + y_i)} \quad (3.15)$$

is a critical one: we have

$$\beta_G = \frac{1-D}{1+D} \quad (3.16)$$

and

$$\beta_E = 1-D \quad (3.17)$$

It is clearly a distance function in that it measures "apartness". However, (3.15) implies that it is not always non-negative when $x, y \notin \mathbb{R}^{n+}$. Although our definitions of the x_i and y_i (equation (3.7)) imply that $x, y \in \mathbb{R}^{n+}$, it would seem useful, at least mathematically, for it to be non-negative in the whole of \mathbb{R}^n .

Two generalisations to provide for this suggest themselves:

$$D_1(x, y) = \frac{\sum_{i=1}^n |x_i - y_i|}{\sum_{i=1}^n [|x_i| + |y_i|]} \quad (3.18)$$

and

$$D_2(x, y) = \frac{\sum_{i=1}^n |x_i - y_i|}{\sum_{i=1}^n |x_i + y_i|} \quad (3.19)$$

Both reduce to (3.15) when $x, y \in \mathbb{R}^{n+}$.

But (3.18) and (3.19) suggest a further generalisation, along the same lines as the Minkowski distances: taking $\beta > 0$,

$$D_{1\beta}(x,y) = \frac{(\sum_{i=1}^n |x_i - y_i|^\beta)^{1/\beta}}{(\sum_{i=1}^n |x_i|^\beta)^{1/\beta} + (\sum_{i=1}^n |y_i|^\beta)^{1/\beta}} \quad (3.20)$$

$$\text{and } D_{2\beta}(x,y) = \frac{(\sum_{i=1}^n |x_i - y_i|^\beta)^{1/\beta}}{(\sum_{i=1}^n |x_i + y_i|^\beta)^{1/\beta}} \quad (3.21)$$

respectively. Again, both reduce to (3.15) when $x, y \in \mathbb{R}^{n+}$ and $\beta = 1$.

These are our "psychological distance functions" for prothetic continua.

3.1.2 Discussion

Fuller discussion of the properties of these functions can be found in Chapter 4, but a few comments are in order here.

Firstly, what do the x_i represent? Originally (equation (3.8)) they were introduced merely as notational simplifications. However, if Restle's theory has any basis in reality, what we called the x_i must be measurable. His theory in fact says that prothetic continua (which we assumed in assumption P.(b)) are

intensive or quantitative continua; the x_i should represent the subjective magnitude of the stimulus x on the i th continuum or dimension. Further, since $x_i = 0$ at zero intensity on the i th dimension, negative values of x_i now have a special meaning - if they are permissible: they may not be. This also implies that the x_i -scales are at least ratio scales.

Secondly, what does the parameter β mean? There is no obvious interpretation, but mathematically it represents an increasing merging of the separate dimensions - perhaps in the sense that they increasingly (but independently) "cross common ground". Initially, for $\beta = 1$, we assumed (P.(c)) that the dimensions were totally disjoint; larger β may in some way represent larger intersections between the dimensions - though not larger dependence between them. Alternatively, it may represent changes in the form of the measure, \underline{m} . Empirically, we may find it has a similar interpretation (in terms of analyzability of stimuli) to the Minkowski-distance parameter. Note however, that we do not yet restrict β to being greater than or equal to unity, as is generally true in the Minkowski case.

Thirdly, the general form of $D_{1\beta}$ and $D_{2\beta}$, at least for $\beta \geq 1$, is of a normalised distance.

We shall often refer to them in this way. However, this also suggests further generalisations of the distance function which may be carried out virtually ad infinitum: take any metric on \mathbb{R}^n and normalise it in some way. The examples given here suggest two ways immediately: if d is the metric, then two normalised metrics from d are

$$\frac{d(x,y)}{d(x,\theta)+d(y,\theta)}$$

and

$$\frac{d(x,y)}{d(x+y,\theta)}$$

The triangle inequality always means the first is not greater than 1.0; the second is not necessarily as well behaved: for example, if $x = -y$ (if points not in \mathbb{R}^{n+} are permitted) the denominator is zero. The first runs into problems of zero denominator only when $x = y = \theta$, (this follows from the facts that, since d is a metric, it is non-negative, and $d(x,y) = 0 \Rightarrow x = y$); we shall define either normalised metric to be zero in this case.

Finally, the distance function can be related to other content models. By derivation, we have (at least for $x,y \in \mathbb{R}^{n+}$)

$$g_G = \frac{1-D_{11}}{1+D_{11}}$$

Goude's (1966) model (see table 2.1) is

$$g_{GD}(x,y) = 1 - \frac{\sqrt{h_x^2 + h_y^2 - 2h_x h_y \cos \phi_{xy}}}{\sqrt{h_x^2 + h_y^2 + 2h_x h_y \cos \phi_{xy}}}$$

By elementary Euclidean geometry, this can be rewritten as

$$g_{GD}(x,y) = 1 - \frac{\left(\sum_{i=1}^n (x_i - y_i)^2 \right)^{\frac{1}{2}}}{\left(\sum_{i=1}^n (x_i + y_i)^2 \right)^{\frac{1}{2}}}$$

$$\text{so } g_{GD} = 1 - D_{22} \quad (3.22)$$

Lastly, and, mathematically, most interestingly, Micko's interpretation of the vector content models gives, in the most general form,

$$g_M(x,y) = 1 - \frac{d(x,y)}{d(x,e) + d(y,e)}$$

(Micko, 1970, p. 220: see section 2.1.1.1), where d is any distance on \mathbb{R}^n . Thus as a special case, we have

$$g_M = 1 - D_{1\beta} \quad (3.23)$$

for any $\beta \geq 1$; otherwise Micko's models form a

subset of the generalised "normalised metrics" suggested in the previous paragraph. Thus we have several tie-ups between the present distance functions and some content models which have already been partially investigated empirically. It is also of interest that D_{11} is the Bray-Curtis measure (see section 2.1.3.6).

3.2 THE MODEL ON METATHETIC CONTINUA

We shall look at the two models (3.1) and (3.2) on metathetic continua of a special kind - those in which all member sets are of equal measure. It is by no means a general case, but it is at the opposite end of the metatheticness - protheticness continuum from protheticness, easily produces interesting and suggestive results, and is reasonably plausible. Other metathetic continua will be briefly discussed later.

3.2.1 The Model

Consider a space B of "relevant aspects", constructed as follows:

For $i = 1, 2, \dots, n$, let

$$\begin{aligned} B_i' &= \{B_{ia}' : a \in \Omega_i\} \\ B_i'' &= \{B_{ia}'' : a \in \Omega_i\} \end{aligned} \quad (3.24)$$

where

$$M.(a) : \Omega_i \subseteq [0,1]$$

$$M.(b) : \alpha < \alpha' \Rightarrow \left. \begin{array}{l} B'_{i\alpha} \subseteq B'_{i\alpha'} \\ \text{and } B''_{i\alpha'} \subseteq B''_{i\alpha} \end{array} \right\} \quad \forall \alpha, \alpha' \in \Omega_i$$

(i.e. B'_i and B''_i , the "polar arrays", are families of nested sets which are, respectively, increasing and decreasing).

$$M.(c): B'_{i\alpha} \cap B''_{i\alpha'} = \emptyset \quad \forall \alpha, \alpha' \in \Omega_i$$

$$M.(d): \underline{m}(B'_{i\alpha}) = m_i - \underline{m}(B''_{i\alpha}) = \alpha m_i$$

for some $m_i \in \mathbb{R}^+$.

Now define, for $i=1,2,\dots,n$,

$$B_i = \{B_{i\alpha} : B_{i\alpha} = B'_{i\alpha} \cup B''_{i\alpha} ; \alpha \in \Omega_i\} \quad (3.25)$$

and further assume

$$M.(e): B_{i\alpha} \cap B_{j\alpha'} = \emptyset \quad \forall \alpha \in \Omega_i, \alpha' \in \Omega_j,$$

$$\forall i, j \in \{1, 2, \dots, n\}, i \neq j$$

Then, finally, the space itself can be defined:

$$\mathcal{B} = \{X: X = \bigcup_{i=1}^n X_i; \quad X_i \in \mathcal{B}_i, i=1,2,\dots,n\}.$$

Thus, the space, \mathcal{B} , is a family of unions of n disjoint metathetic continua, the \mathcal{B}_i . Each set in \mathcal{B} is a unique union of sets from each continuum, so again we have a dimensional structure. Assumption M.(a) serves exactly the same purpose as P.(a); we take the interval $[0,1]$ instead of the whole non-negative real line, \mathbb{R}^+ , solely for reasons of convenience related to M.(d), as will be seen below. Assumption M.(b), along with defining equations (3.24), sets up the two "polar arrays" on each dimension which together define, through (3.25) the metathetic continuum. Assumptions M.(c) and M.(e) together act in the same way as assumption P.(c) to prevent overlap between dimensions.

Assumptions M.(c) and M.(d) make up the "measure equality" assumption, that on each dimension \mathcal{B}_i , all sets have equal measure:

Let $X_i \in \mathcal{B}_i$; then, by (3.25), for some $B'_{i\alpha} \in \mathcal{B}'_i$, $B''_{i\alpha} \in \mathcal{B}''_i$,

$$X_i = B'_{i\alpha} \cup B''_{i\alpha}$$

so

$$\underline{m}(X_i) = \underline{m}(B'_{i\alpha} \cup B''_{i\alpha})$$

$$= \underline{m}(B'_{i\alpha}) + \underline{m}(B''_{i\alpha}) \text{ by M.(c) and}$$

since \underline{m} is a

measure.

$$= \alpha m_i + (1-\alpha)m_i \text{ by assumption}$$

M.(d).

$$= m_i$$

$$\text{i.e. } \forall X_i \in \mathcal{B}_i, \quad \underline{m}(X_i) = m_i \quad (3.26)$$

This means that the "size" (i.e. measure) of a set on one of the metathetic dimensions making up the space is not a measure of its position on that dimension.

In spite of the space being "metathetic", the basic structure is still the same as in the prothetic case: any set in \mathcal{B} can be expressed as a unique union of sets from $2n$ prothetic continua:

$$X = \bigcup_{i=1}^{2n} X_i, \text{ where, for some } x_i \in \Omega_i \text{ if } i \leq n \quad (3.27)$$

$$X_i = B'_{ix_i} \in \mathcal{B}'_i \text{ and } X_{n+i} = B''_{ix_i} \in \mathcal{B}''_i$$

Note that for each X , each x_i is unique by a unique union property proved in section 3.1.1.

Thus the families B_i' and B_i'' are the disjoint prothetic arrays corresponding to the α_i of section 3.1.1. We can therefore go through an identical process to that section to get equation (3.15) in the form

$$D(x,y) = \frac{\sum_{i=1}^{2n} |\underline{m}(X_i) - \underline{m}(Y_i)|}{\sum_{i=1}^{2n} (\underline{m}(X_i) + \underline{m}(Y_i))} \quad (3.28)$$

Now, by (3.27) and M.(d) we have, for $i \leq n$

$$\underline{m}(X_i) = m_i - \underline{m}(X_{n+i}) = x_i m_i$$

and
$$\underline{m}(Y_i) = m_i - \underline{m}(Y_{n+i}) = y_i m_i$$

Thus

$$|\underline{m}(X_i) - \underline{m}(Y_i)| = |\underline{m}(X_{n+i}) - \underline{m}(Y_{n+i})| = m_i |x_i - y_i|$$

and

$$\underline{m}(X_i) + \underline{m}(X_{n+i}) = \underline{m}(Y_i) + \underline{m}(Y_{n+i}) = m_i$$

Therefore (3.28) becomes

$$D(x,y) = \frac{2 \sum_{i=1}^n m_i |x_i - y_i|}{2 \sum_{i=1}^n m_i}$$

$$= \sum_{i=1}^n \frac{m_i}{\sum_{j=1}^n m_j} |x_i - y_i|$$

and putting $n_i = m_i / \sum_{j=1}^n m_j$ for $i=1, 2, \dots, n$,

$$D(x,y) = \sum_{i=1}^n n_i |x_i - y_i| \quad (3.29)$$

$$\text{where } \sum_{i=1}^n n_i = 1.$$

Thus, the psychological distance function for the metathetic case is the (weighted) City-Block distance. Compare also the distance of Gower (1967) - see section 2.1.3.6, The similarity models for metathetic continua therefore are (by equations (3.16) and (3.17) respectively)

$$g_G(x,y) = \frac{1 - \sum_{i=1}^n n_i |x_i - y_i|}{1 + \sum_{i=1}^n n_i |x_i - y_i|} \quad (3.30)$$

$$g_E(x,y) = 1 - \sum_{i=1}^n n_i |x_i - y_i| \quad (3.31)$$

The obvious generalisation of (3.29), following equation (3.20) more than equation (3.21), is to the Minkowski distances - call them M_β - although we may possibly also take $0 < \beta < 1$. Thus the Minkowski models are theoretically reasonable "psychological distance functions" for equal-measure metathetic continua. The fact that they are a special case of the $D_{1\beta}$ (or $D_{2\beta}$) models suggests, however, that the latter are the more basic models in this area of psychology.

It should be remembered that the x_i do not have the same meaning here as the x_i in the general model of section 3.1.1. By equation (3.27), in conjunction with M.(d), x_i is the proportion the increasing* polar array set, B'_{ix_i} , is of the total set $X (= B_{ix_i})$ on the metathetic continuum. It also in a sense represents "how far" the set is along that continuum (see Figure 3.1). In the metathetic case, x_i does not represent the measure of the set X_i as it does in the prothetic case. For any metathetic continuum in fact, a

* Which of the polar arrays is to be taken as "increasing" is largely arbitrary, but must somehow be fixed.

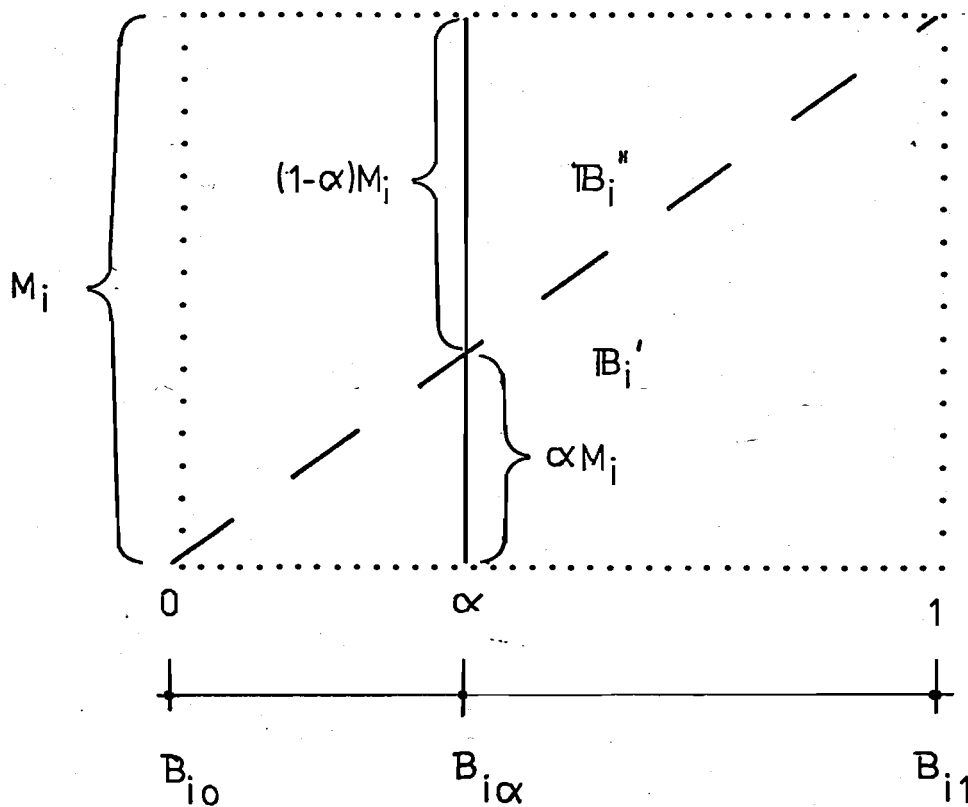


FIGURE 3.1: The equal-measure metathetic continuum.

Each set $B_{i\alpha}$ on the continuum is a disjoint union of two sets: one, of size αm_i , from the (increasing) polar array TB_i' , the other, of size $(1-\alpha)m_i$, from the (decreasing) polar array TB_i'' .

$$\underline{m}(B_{i\alpha}) = \underline{m}(TB_i' \cup TB_i'') = m_i$$

If $x_i = B_{i\alpha}$, we write $\alpha = x_i$.

relation analogous to $M.(d)$ will give the measure of the total set X_i from the value x_i . Also in this metathetic case the x_i are bounded to unit range, thus bounding $D (= \mathbb{M}_\beta)$; this is essential if equations (3.30) and (3.31) are not to be negative, and has wide implications as to permissible generalisation functions (see section 2.1.2.1).

3.2.2 Discussion

The first question that comes to mind about this result is: what, empirically, is an "equal-measure" metathetic continuum? The equal-measure property is very reminiscent of the equal-intensity (or equal vector length) assumption used to simplify vector models (see section 2.1.1.2). While the two cases may not be identical (for example, since all the vector content models are based in Euclidean space, $M.(d)$ would probably be replaced with

$$\underline{m}(B'_{i\alpha}) = \sqrt{m_i^2 - [\underline{m}(B''_{i\alpha})]^2} = \alpha m_i),$$

the same intuitions are involved. The continua used to test the vector models under the equal

intensity assumption are therefore examples that could be used here, at least approximately. Some are given in section 2.1.2.2; another would seem to be the Conservatism-Liberalism scale in Lund (1974) (see also Stevens, 1966). This equal-measure restriction does not seem too unrealistic.

To use Lund's continuum as an example, the scale would be made up of two, presumably prothetic, continua: "Conservatism" and "Liberalism". Each point on the scale would be of equal "intensity", consisting of a certain proportion of Liberalism, and a certain proportion of Conservatism, one of which we measure to give the scale value. Of course, the interpretation of this example - on a purely subjective continuum - may not be typical of psychophysical scales whose metatheticness or otherwise may be determined largely by physiological factors, and which therefore require a physiological interpretation (e.g. pitch: see Stevens and Galanter, 1957).

A large part of the construction of the previous section is only a matter of definition. As long as they are bounded with unit range, there is no reason why the x_i -scales should

directly measure proportionality. Since only their differences are actually used (equations (3.30) and (3.31)), they can therefore be translated by any amount. In practice, of course, the x_i -scales can also all be multiplied by any positive amount - so they are interval scales - but we shall use the convention that any such multiplication will be absorbed into the weights $\{m_i\}_{i=1}^n$ or $\{n_i\}_{i=1}^n$.

The status of these weights is obviously dependent on that of the measure \underline{m} . In so far as \underline{m} varies between observers, as it may well do, the weights obviously may change. Thus the weights have a similar interpretation to that given by Carroll (1972) or Gregson (1972), for example. The measure may of course vary in other ways: see section 3.3.

The present case of metathetic continuum is a very special one, although it may in fact be empirically quite common, or at least a good approximation to many others. Using the line of reasoning given here, it is obviously possible to derive results for any such metathetic continuum - at least in Restle's sense: see Restle (1961, p. 50 ff.) for examples. Their forms are not in

general as straightforward as the one given here, but they may be found useful at some stage. Eisler (1963) in fact suggested that such results were necessary, suggesting that "protheticness" should be quantifiable. Although he gave no such quantification, Restle's analysis immediately suggests some: for example, using the notation of this chapter, the index

$$\int_{\Omega} \lim_{\Delta\alpha \rightarrow 0} \left\{ \frac{\min(\underline{m}(X_{\alpha} - X_{\alpha+\Delta\alpha}), \underline{m}(X_{\alpha+\Delta\alpha} - X_{\alpha}))}{\max(\underline{m}(X_{\alpha} - X_{\alpha+\Delta\alpha}), \underline{m}(X_{\alpha+\Delta\alpha} - X_{\alpha}))} \right\} d\alpha \bigg/ \int_{\Omega} d\alpha$$

for a general metathetic array $\{X_{\alpha} : \alpha \in \Omega\}$ is zero for prothetic continua and unity for the equal-measure metathetic continua defined in section 3.2.1.

3.3 OTHER COMMENTS

Obviously, certain aspects will be common to the prothetic and metathetic cases. Firstly, both have been given a dimensional structure. It is obvious from the derivations that this is by no means necessary, though very convenient. It is possible to generalise the models to non-dimensional and/or discrete (discontinuous) spaces. Whether this can be done usefully is a different matter;

we shall not investigate it here. One should always bear in mind, however, that the interpretation of the dimensions in the two cases is different - see section 3.2.1.

Secondly, the dimensions are based on Restle's theory of arrays. Restle's arrays always have a "core" which would represent, in our terms, irrelevant aspects. By considering only the space of relevant aspects, we avoid the problem of having to consider the core, but it may still need to be taken into account in some situations. It may particularly be relevant to context effects. Being common to all stimuli but disjoint from their relevant aspects, its effect would be to add the same amount to both numerator and denominator of all our models. (ref. equations (3.6) and (3.7)).

Thirdly, the whole structure is based on the form of the measure function \underline{m} . Its properties must have an all-pervasive effect on the construction, so they must be closely examined. It is merely a mathematical nicety, for example, that we accept a priori the property

$$X \cap Y = \emptyset \Rightarrow \underline{m}(X \cup Y) = \underline{m}(X) + \underline{m}(Y).$$

Any number of replacements for this could be thought of, the most appealing perhaps being

$$X \cap Y = \emptyset \Rightarrow \underline{m}(X \cup Y) = [(\underline{m}(X))^\beta + (\underline{m}(Y))^\beta]^{1/\beta}$$

for some $\beta \geq 1$. This would give consistently larger results than the corresponding models using $D_{2\beta}$; it would of course coincide with $D_{1\beta}$, $D_{2\beta}$ and M_β when $\beta = 1$.

A final property the two types of model have in common is the form of the similarity gradient: Gregson's model gives the relation

$$\mu_G = \frac{1-D}{1+D} \quad (3.35)$$

and Eisler's

$$\mu_E = 1-D \quad (3.36)$$

(3.35) seems to be a new uninvestigated form for the similarity gradient. We shall look at some of the consequences of these equations in subsequent chapters.

Finally, two comments on the relationship between the two spaces. Firstly, although we draw the distinction between protheticness and metatheticness, as already noted, the former is but a special case of the latter. Equally, metathetic spaces can be thought of as a particular type of prothetic space, as can be seen from the construction of section 3.2.1: a metathetic continuum consists of unions of sets from two prothetic continua. What differentiates the two types of space is the degree of correlation (in a loose sense) between pairs of prothetic continua. A metathetic continuum is in a sense a pair of highly (negatively) correlated prothetic continua. In intermediate cases it may be difficult to say whether the space is prothetic or metathetic.

This leads to the second comment: we have dealt in our models only with spaces consisting only of prothetic continua or only of metathetic continua. In real life, it is more likely that some of the dimensions will be of one type and some of the other. For example, when comparing consumer goods, one compares them both on their prothetic attributes (e.g. price, life expectancy, weight), and on their attributes which are probably metathetic (e.g. "quality", convenience,

aestheticness). In such a case, we would have a mixed model, such as

$$S_G(x, y) = \frac{\sum_{i=1}^{n_1} m_i (1 - |x_i - y_i|) + \sum_{j=n_1+1}^{n_2} \min(x_j, y_j)}{\sum_{i=1}^{n_1} m_i (1 + |x_i - y_i|) + \sum_{j=n_1+1}^{n_2} \max(x_j, y_j)}$$

for Gregson's model, where the x_i and y_i refer to equal-intensity metathetic continua for $i=1, 2, \dots, n_1$, and to prothetic continua for $i=n_1+1, \dots, n_2$. In research, experimenters, perhaps subconsciously, often seem to try to avoid such situations by instruction or by choice of stimuli. It should nevertheless be borne in mind.

CHAPTER 4

PROPERTIES OF THE PSYCHOLOGICAL DISTANCE FUNCTION.

In this chapter, an initial investigation into the properties of the $D_{1\beta}$ and $D_{2\beta}$ distance functions will be presented. If these functions are found to have any relevance, a far more thorough investigation would be in order. The one given here does not pretend to be general: it is aimed mainly at putting these functions and the Minkowski distances, M_β , into perspective, and to provide a rough picture of their properties so that some intuitions about them can be built up.

4.1 THE MULTIDIMENSIONAL CASE

4.1.1 Properties discussed in Chapter 1

We shall first see how $D_{1\beta}$ and $D_{2\beta}$, and the corresponding similarity models for S_G and S_E , fare with respect to the conditions D1-D8 and S1-S8. In each case we shall state the relevant condition and then, if necessary, prove or disprove it for each case and discuss the result. Some of the more technical proofs from this chapter have been put in Appendix A.

4.1.1.1 Properties D1-D8

$$D1: 0 \leq D(x,y) < \infty \quad \forall x,y \in \mathbb{R}^n$$

This is clearly true for both $D_{1\beta}$ and $D_{2\beta}$.

In fact, for $x,y \in \mathbb{R}^{n+}$ (actually, when x and y are in the same "quadrant"*) we have that

$$0 \leq D(x,y) \leq 1.$$

In the case of $D_{1\beta}$ this is true because of the Minkowski Inequality (e.g. Beckenbach and Bellman, 1971, p. 25) and is true for all $x,y \in \mathbb{R}^n$, $x,y \neq e$; in the case of $D_{2\beta}$ it is obvious. However this breaks down

(a) when $x = y = e$, in which case we define

$$D_{1\beta}(e,e) = D_{2\beta}(e,e) = 0 \quad (4.1)$$

and (b) for $D_{2\beta}$ when x and y are not in the same "quadrant". In this second case, $D_{2\beta}$ may be greater than unity, and in the most extreme case, when $x = -y$, its denominator is zero, so that $D_{2\beta}$ is undefined. It is difficult to interpret this psychologically especially since unit is the upper bound ("total dissimilarity") in \mathbb{R}^{n+} .

*We shall use "quadrant" (in quotes) to mean the n -dimensional counterpart to the quadrant in two dimensions: i.e. those parts of \mathbb{R}^n in which, for $1 \leq i \leq n$, all i th coordinates have the same sign.

An important reason for wanting the unity upper bound is so that similarity is non-negative when calculated from \mathbb{D} by either similarity function (3.35) or (3.36). (This applies equally to the Minkowski model of section 3.2.1). Thus the unboundedness of $\mathbb{D}_{2\beta}$ may be an argument against it - although its bad characteristics only appear for points not in \mathbb{R}^{n+} , which we may not need.

$$D2: \quad \mathbb{D}(x, x) = 0 \quad \forall x \in \mathbb{R}^n$$

This is clearly true for all $x \neq e$; our definition (4.1) completes the condition.

$$D3: \quad \mathbb{D}(x, y) = 0 \Rightarrow x = y$$

It is clear that this condition is true for both $\mathbb{D}_{1\beta}$ and $\mathbb{D}_{2\beta}$ since if either distance function is zero, we must have

$$\left(\sum_{i=1}^n |x_i - y_i|^\beta \right)^{1/\beta} = 0,$$

which implies that $x = y$. Note though that we can find arbitrarily small values of $\mathbb{D}_{1\beta}(x, y)$ and $\mathbb{D}_{2\beta}(x, y)$ by taking x and y sufficiently far from the origin, keeping $(\sum |x_i - y_i|^\beta)^{1/\beta}$ constant (see comments on D5 below).

$$D4: \quad \mathbb{D}(x, y) = \mathbb{D}(y, x) \quad \forall x, y \in \mathbb{R}^n$$

This is obvious.

D5: $D(x + a, y + a) = D(x, y) \quad \forall x, y, a \in \mathbb{R}^n.$

Clearly, this is not generally true for $D_{1\beta}$ or

$D_{2\beta}$. In fact, if x, y, a are in the same "quadrant" (e.g. $x, y, a \in \mathbb{R}^{n+}$), it is easily seen that, if $a \neq 0$,

$$\left(\sum_{i=1}^n |x_i + a_i|^\beta \right)^{1/\beta} > \left(\sum_{i=1}^n |x_i|^\beta \right)^{1/\beta}$$

(and similarly with y)

$$\text{and } \left(\sum_{i=1}^n |(x_i + a_i) + (y_i + a_i)|^\beta \right)^{1/\beta} > \left(\sum_{i=1}^n |x_i + y_i|^\beta \right)^{1/\beta}$$

so that, for all x, y, a in the same "quadrant",

$a \neq 0$,

$$\left. \begin{aligned} D_{1\beta}(x + a, y + a) &< D_{1\beta}(x, y) \\ D_{2\beta}(x + a, y + a) &< D_{2\beta}(x, y) \end{aligned} \right\} \quad (4.2)$$

(A more general case for $D_{2\beta}$ is when a is in the same "quadrant" as $x+y$. More general, but less easily stated, conditions exist.)

This is directly compatible with S5, and, as discussed in section 1.5.2, is psychologically more reasonable in the prothetic case than D5. It also implies that the coordinates of x and y cannot be on interval scales for consistent similarity judgments.

D6: $D(\alpha x, \alpha y) = |\alpha| D(x, y) \quad \forall x, y \in \mathbb{R}^n, \alpha \in \mathbb{R}.$

This is not true of $D_{1\beta}$ or $D_{2\beta}$. We show it for $D_{1\beta}$; the proof for $D_{2\beta}$ is virtually identical:

$$\begin{aligned} D_{1\beta}(\alpha x, \alpha y) &= \frac{(\sum_{i=1}^n |\alpha x_i - \alpha y_i|^\beta)^{1/\beta}}{(\sum_{i=1}^n |\alpha x_i|^\beta)^{1/\beta} + (\sum_{i=1}^n |\alpha y_i|^\beta)^{1/\beta}} \\ &= \frac{|\alpha| (\sum_{i=1}^n |x_i - y_i|^\beta)^{1/\beta}}{|\alpha| (\sum_{i=1}^n |x_i|^\beta)^{1/\beta} + (\sum_{i=1}^n |y_i|^\beta)^{1/\beta}} \\ &= D_{1\beta}(x, y) \end{aligned}$$

Again, this is compatible with condition S6, which was seen to be more reasonable psychologically, in section 1.5.3. It also suggests that the coordinates of x and y are on a ratio scale.

D7: $D(x, y) \leq D(x, z) + D(z, y) \quad \forall x, y, z \in \mathbb{R}^n.$

For almost every β , counterexamples to D7, the triangle inequality, can easily be found. The general pattern of violations seems to be as follows:*

*These comments, unless otherwise stated, are derived from a computer investigation of violations of the triangle inequality and have yet to be rigorously proved or disproved.

- a) Comparing models, with β constant, for $\beta \geq 1$, the difference $D(x,y) - D(x,z) - D(z,y)$ is smaller for $D_{1\beta}$ than for $D_{2\beta}$. Thus there are less violations of the triangle inequality for $D_{1\beta}$ than for $D_{2\beta}$, and violations, when they occur, are not as great for $D_{1\beta}$. For $0 < \beta < 1$, the situation is not as simple, but the reverse often applies.
- b) Varying β , the greatest violations as measured by $D(x,y) - D(x,z) - D(z,y)$ occur when $\beta < 1$ (recall that for $\beta < 1$, the Minkowski distance functions, M_β , also disobey the triangle inequality). For $1.5 < \beta < 3$, violations are at their smallest. In fact for $D_{1\beta}$ and $\beta \neq 2$, violations are quite minor and difficult to find; for D_{12} , no violations have yet been found. Violations are again larger for large β greater than 3, but they do not reach the same magnitude as for β near zero.
- c) The worst violations occur when two of the points in a triple are on different axes or in opposite "quadrants".

On the whole, except for $D_{1\beta}$ with $\beta > 1.5$, violations can be quite large (to the

extent that they would probably be statistically significant in an experimental situation represented by these models). But in an experimental situation, they may not be very ubiquitous if the experimental design does not provide many points near different axes or in opposite "quadrants" since violations not of this type generally give $D(x,y) - D(x,z) - D(z,y)$ under 10% of $D(x,y)$. For $D_{1\beta}$, $\beta \geq 1.5$, virtually all violations are, in this sense, well below 10%.

A final important point here is that for $\beta \geq 1$, $D_{1\beta}$ obeys the triangle inequality for triples on straight lines (colinear triples). This is proved in Appendix A. It is not generally true of $D_{2\beta}$ - counterexamples are easily found - although it is of course true for D_{21} for points in \mathbb{R}^{n+} , since then $D_{21} = D_{11}$. The special case of points on lines through the origin is dealt with in more detail in section 4.2.

$$D8: D(x,y) \geq D(x_i, y_i), \quad \forall i = 1, \dots, n, \quad \forall x, y \in \mathbb{R}^n.$$

(Recall that x_i was defined rather untidily as

$(0, 0, \dots, 0, x_i, \dots, 0) \in \mathbb{R}^n$, where

$x = (x_1, x_2, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) \in \mathbb{R}^n$.) Both

examples and counterexamples to D8 are easily found (e.g. $x = (2,4)$, $y = (6,5)$), so it does not hold in general. However, as remarked in Chapter 1, it can be made to hold under certain conditions if D8 is redefined. The line of reasoning here is that D8 expresses the idea that overall dissimilarity between two stimuli is always greater than the dissimilarity between any two stimuli which, while having, respectively, the same values as the original pair on one attribute, differ only on that attribute. However this is only meaningful if either the values of the other "constant" attributes are specified, or these values are irrelevant. Because of the intradimensional subtractivity of Minkowski distances, the latter of these meaningfulness conditions applies, so the values of the constant attributes do not need to be specified: they are usually set at zero, as in the definition of x_i above. In the case of $D_{1\beta}$ and $D_{2\beta}$, however, this condition does not apply, so one must specify the constant values. D8 then must be conditional on where these constant values are in the space. We shall do this by reformulating D8 as follows:

Let $x, y \in \mathbb{R}^n$; let $i \in \{1, 2, \dots, n\}$. Let $a(i) \in \mathbb{R}^n$ be such that $a(i)_i = 0$. Then for which $a(i)$ is it true that

$$D(x_i + a(i), y_i + a(i)) \leq D(x, y)? \quad (4.4)$$

For the Minkowski distances, the answer to this question is: for all $a(i) \in \mathbb{R}^n$. For $D_{1\beta}$ and $D_{2\beta}$, answers of varying rigour are available, but the simplest conditions are, respectively,

$$\sum_{j \neq i} |a(i)_j|^\beta \geq \max \left\{ \sum_{j \neq i} |x_j|^\beta, \sum_{j \neq i} |y_j|^\beta \right\} \quad (4.5)$$

and

$$\sum_{j \neq i} |a(i)_j|^\beta \geq \sum_{j \neq i} \left| \frac{x_j + y_j}{2} \right|^\beta \quad (4.6)$$

These conditions are sufficient but not necessary. In addition, for D_{11} we have the stronger (but still not necessary) condition

$$\sum_{j \neq i} |a(i)_j| \geq \min \left\{ \sum_{j \neq i} |x_j|, \sum_{j \neq i} |y_j| \right\} \quad (4.7)$$

The proofs of conditions (4.5) and (4.6) are quite trivial; the proof of (4.7) is given in Appendix A.

4.1.1.2 Properties S1 - S8

We now turn to the corresponding similarity models for \mathcal{S}_G and \mathcal{S}_E . It will be recalled from Chapter 3 that there are at least two possible

transformations of the dissimilarity model which give plausible similarity models (see equations (3.35) and (3.36)): ¹³⁷

$$F_1(x) = 1 - x \quad 0 \leq x \leq 1 \quad (4.8)$$

$$\text{and } F_2(x) = \frac{1 - x}{1 + x} \quad 0 \leq x \leq 1 \quad (4.9)$$

Both of these "similarity gradients" are non-negative, strictly decreasing, and bounded by unity in the interval in which they are defined here ($x : 0 \leq x \leq 1$). From this it follows that many of the properties of $D_{1\beta}$ and $D_{2\beta}$ are paralleled by corresponding properties of β_E and β_G , where

$$\beta_E = F_1(D_{1\beta}) \quad \text{or} \quad F_1(D_{2\beta}) \quad (4.10, 4.11 \text{ resp.})$$

$$\beta_G = F_2(D_{1\beta}) \quad \text{or} \quad F_2(D_{2\beta}) \quad (4.12, 4.13 \text{ resp.})$$

Note that, when they were originally stated in Chapter 1, the properties S1 - S8 were stated only for points in \mathbb{R}^{n+} . In the discussion following, it will be seen that many hold, for β_E and β_G , for all of \mathbb{R}^n .

$$S1: 0 \leq \beta(x,y) \leq 1 \quad \forall x,y \in \mathbb{R}^{n+}.$$

The abovementioned properties of F_1 and F_2 make $S1$ true for $D_{1\beta}$, for all $x,y \in \mathbb{R}^n$, since we have that $0 \leq D_{1\beta} \leq 1$. As remarked in the comments on $D1$, however, $D_{2\beta}$ may be greater than unity, in which case we would have $\beta(x,y) < 0$. Thus for $S1$ to hold for $F_1(D_{2\beta})$ or $F_2(D_{2\beta})$, x and y must be restricted so that $D_{2\beta}(x,y) \leq 1$. One such restriction is $x,y \in \mathbb{R}^{n+}$.

$$S2: \beta(x,x) = 1 \quad \forall x \in \mathbb{R}^{n+}.$$

Transformations F_1 and F_2 along with property $D2$ give this immediately, for all $x \in \mathbb{R}^n$.

$$S3: \beta(x,y) = 1 \Rightarrow x = y, \quad \forall x,y \in \mathbb{R}^{n+}$$

Since both F_1 and F_2 are strictly decreasing, and hence one-to-one, $S3$ follows from $D3$ for all $x,y \in \mathbb{R}^n$. The comments given in the case of $D3$ also apply here, however.

$$S4: \beta(x,y) = \beta(y,x) \quad \forall x,y \in \mathbb{R}^{n+}.$$

Again, this follows directly from $D4$, and is true for all $x,y \in \mathbb{R}^n$.

$$S5: \beta(x+a,y+a) > \beta(x,y) \quad \forall x,y,a \in \mathbb{R}^{n+}, \quad a \neq 0.$$

It was shown under D5 that if x, y, a are in the same "quadrant", $a \neq 0$, then for both $D_{1\beta}$ and $D_{2\beta}$,

$$D(x+a, y+a) < D(x, y).$$

Since F_1 and F_2 are strictly decreasing, S5 follows as the special case when the "quadrant" is \mathbb{R}^{n+} . The comments given in the discussion of D5 also apply here.

$$S6: \quad \beta(ax, ay) = \beta(x, y) \quad \forall x, y \in \mathbb{R}^{n+}, \quad a \in \mathbb{R}^+.$$

It was shown under D6, for both $D_{1\beta}$ and $D_{2\beta}$ that $\forall x, y \in \mathbb{R}^n, \quad a \in \mathbb{R}$,

$$D(ax, ay) = D(x, y).$$

Since, either β is a function of one of $D_{1\beta}$ or $D_{2\beta}$, S6 follows, for all $x, y \in \mathbb{R}^n, \quad a \in \mathbb{R}$.

S7: $\beta(x, y) = 0$ if $x, y \in \mathbb{R}^{n+}$ are on different axes.

$\beta(x, y) = 0$ is clearly equivalent, under transformations F_1 and F_2 , to $D(x, y) = 1$. Using our previous definition of x_i, y_j (see D8), let

$x = x_i, y = y_j, i \neq j$ (i.e. x and y are on different axes in \mathbb{R}^n). Then

$$\mathbb{D}_{1\beta}(x,y) = \frac{(|x_i|^\beta + |y_j|^\beta)^{1/\beta}}{|x_i| + |y_j|}$$

This is equal to 1 only if $\beta = 1$, unless x or y is the origin. But

$$\mathbb{D}_{2\beta}(x,y) = \frac{(|x_i|^\beta + |y_j|^\beta)^{1/\beta}}{(|x_i|^\beta + |y_j|^\beta)^{1/\beta}} = 1.$$

Thus S7 is true (for all $x, y \in \mathbb{R}^n$ on different axes) for the similarity models (4.11) and (4.13), and for the models

$$\beta_E = 1 - \mathbb{D}_{11}$$

$$\beta_G = \frac{1 - \mathbb{D}_{11}}{1 + \mathbb{D}_{11}}$$

However, a weaker property, corresponding to S7 can be proved for the models corresponding to $\mathbb{D}_{1\beta}$ (i.e. (4.10) and (4.12)). From Hölder's inequality (see Beckenbach and Bellman, 1971, p.19) it is easily seen that

$$\frac{\left(\sum_{i=1}^n a_i^\beta\right)^{1/\beta}}{\sum_{i=1}^n a_i} \geq n^{\frac{1-\beta}{\beta}} \quad \forall a \in \mathbb{R}^{n+}, \beta \geq 1.$$

from which it follows that

$$D_{1\beta}(x, y) = \frac{(|x_i|^\beta + |y_j|^\beta)^{1/\beta}}{|x_i| + |y_j|} \geq 2^{\frac{1-\beta}{\beta}}, \beta \geq 1$$

for any x, y defined as above. The number $2^{(1-\beta)/\beta}$ equals 1 when $\beta = 1$ and is always greater than 0.5 when $1 \leq \beta < \infty$. This puts a lower bound on dissimilarities between points on different axes, distinctly differentiating the $D_{1\beta}$ space from, say, a Minkowski space where points on different axes can be arbitrarily close together.

The corresponding properties for models (4.10) and (4.12) are respectively, for x and y on different axes, and $\beta \geq 1$,

$$\beta_E(x, y) \leq 1 - 2^{(1-\beta)/\beta}$$

and

$$\beta_G(x, y) \leq \frac{1 - 2^{(1-\beta)/\beta}}{1 + 2^{(1-\beta)/\beta}}$$

Note that

$$1 - 2^{(1-\beta)/\beta} \leq \frac{1}{2} \quad 1 \leq \beta < \infty,$$

$$\text{and } \frac{1 - 2^{(1-\beta)/\beta}}{1 + 2^{(1-\beta)/\beta}} \leq \frac{1}{3} \quad 1 \leq \beta < \infty.$$

Thus S7 holds (in all of \mathbb{R}^n) in letter for models (4.11) and (4.13), and in spirit for (4.10) and (4.12) for $\beta \geq 1$.

$$S8: \quad \mathcal{J}(\mathbf{e}, \mathbf{x}) = 0 \quad \forall \mathbf{x} \in \mathbb{R}^{n+}$$

$$\text{Since } D_{1\beta}(\mathbf{e}, \mathbf{x}) = \frac{(\sum |x_i|^\beta)^{1/\beta}}{(\sum |x_i|^\beta)^{1/\beta} + 0} = 1$$

and

$$D_{2\beta}(\mathbf{e}, \mathbf{x}) = \frac{(\sum |x_i - 0|^\beta)^{1/\beta}}{(\sum |x_i + 0|^\beta)^{1/\beta}} = 1$$

and $F_1(1) = F_2(1) = 0$, we have S8 for all $\mathbf{x} \in \mathbb{R}^n$.

4.1.1.3 Other Properties from Chapter 1

From knowledge of the relationship between \mathcal{J} and D , it is clear that all our \mathcal{J} models have the property corresponding to the revised condition D8: i.e.

$$\mathcal{J}(\mathbf{x}, \mathbf{y}) \leq \mathcal{J}(\mathbf{x}_i, \mathbf{y}_i) \quad , \quad i = 1, \dots, n$$

under the conditions discussed. Equally, both D models have the properties corresponding to S7 and S8:

$$\begin{aligned} D_{2\beta}(x,y) &= 1 \quad \text{if } x,y \in \mathbb{R}^n \text{ are on different axes,} \\ D_{1\beta}(x,y) &\geq 2^{(1-\beta)/\beta} \quad \text{if } x,y \in \mathbb{R}^n \text{ are on different axes,} \\ \beta &\geq 1; \\ \text{and } D(e,x) &= 1 \quad \forall x \in \mathbb{R}^n. \end{aligned}$$

No obvious similarity counterpart to D7, the triangle inequality, suggests itself; the inequality

$$\beta(x,y) \geq \beta(x,z)\beta(z,y) \quad \forall x,y,z \in \mathbb{R}^n \quad (4.14)$$

mentioned in Chapter 1, does not hold for any of the four models: counterexamples are easily found. However, since this inequality, which we shall call the "similarity triangle inequality", is the only one so far seriously suggested as a desirable one for similarity to obey, a little more investigation into its relationship to the distance triangle inequality, D7, seems appropriate.

It turns out that this relationship is heavily

dependent on the properties of the functions F_1 or F_2 (equations (4.8), (4.9)). It is easily seen for both F_1 and F_2 that

$$F(a+b) \leq F(a)F(b) \quad \forall a, b, 0 \leq a, b \leq 1 \quad (4.15)$$

with equality only when $a = 0$ or $b = 0$. Now suppose that we have a triple (x, y, z) , $x, y, z \in \mathbb{R}^n$, which under the distance function D disobeys the triangle inequality D7 so that say,

$$D(x, y) > D(x, z) + D(z, y).$$

To ensure that \mathcal{F} is well defined (namely, non-negative: see equations (4.10) - (4.13)) we shall not allow points that give $D(x, y) > 1$ (as may occur with $D_{2\beta}$, or for $D_{1\beta}$ with $\beta < 1$). We shall also assume that $z \neq x$ and $z \neq y$ so that $D(x, z) \neq 0$ and $D(z, y) \neq 0$. Thus $0 < D(x, z) + D(z, y) \leq 1$ and since $F_1(a)$ and $F_2(a)$ are strictly decreasing functions for $0 \leq a \leq 1$, we have

$$\begin{aligned} F(D(x, y)) &\leq F(D(x, z) + D(z, y)) \\ &< F(D(x, z))F(D(z, y)) \quad (\text{using (4.15)}) \end{aligned}$$

for either F_1 or F_2 . Thus

$$\mathcal{F}(x, y) < \mathcal{F}(x, z)\mathcal{F}(z, y)$$

(from equations (4.10) - (4.13)). So any violation

of the triangle inequality for the distance functions immediately implies a violation of the similarity triangle inequality for our similarity models. These are not the only violations of the similarity triangle inequality: enough additional ones exist that similarity models based on D_{12} still have violations.

One might note here that it can even be seen that F_1 and F_2 applied to the Minkowski distances as in Chapter 3, lead to violations of the similarity triangle inequality. For example, if x, y, z are distinct points lying on a straight line so that, say, $M_\beta(x, z) = M_\beta(x, y) + M_\beta(y, z)$, then since equality occurs in (4.15) only when $a = 0$ or $b = 0$, we have

$$F(M_\beta(x, z)) < F(M_\beta(x, y))F(M_\beta(y, z))$$

(recall that in Chapter 3 the distances were restricted to being not greater than unity) so that a violation of the similarity triangle inequality occurs on straight lines.

For the model $F_2(D_{12})$, the similarity triangle inequality does have a special significance however:

on straight lines, similarities are multiplicative - that is, if y lies between x and z on a straight line in \mathbb{R}^n then, if $\beta_G = F_2(D_{12})$,

$$\beta_G(x, z) = \beta_G(x, y)\beta_G(y, z) \quad (4.16).$$

The proof of this is given in Appendix A; it is also true for the one dimensional case of the $F_2(D_{1\beta})$ models: see section 4.2.

This leads naturally to a short discussion of the conditions for equality of the triangle inequality D7. Firstly, the two distance functions do not have segmental additivity: that is, for a pair of points $x, y \in \mathbb{R}^n$, there is not necessarily a point $z \in \mathbb{R}^n$ with $D(x, y) = D(x, z) + D(y, z)$. This means that one of Beals, Krantz and Tversky's (1968, for example) most important axioms does not hold. Secondly, neither of the distance functions are additive on straight lines. Additive triples are generally "nearly" linear, but they can be far enough from linear for assumptions such as those made by Shepard (1957) or Beals and Krantz (1967) to cause significant errors.

The behaviour of the various models under the various conditions of the two triangle inequalities is summarised in Table 4.1.

TABLE 4.1

THE TRIANGLE INEQUALITIES

The behaviour of the Normalised Distance Functions and their corresponding Similarity Functions with respect to various conditions of the two triangle inequalities.

Notation: T.I. - the triangle inequality (D7): $D(x,y) \leq D(x,z) + D(z,y) \forall x,y,z \in \mathbb{R}^n$

S.T.I. - the similarities triangle inequality (4.14): $\beta(x,y) \geq \beta(x,z)\beta(z,y) \forall x,y,z \in \mathbb{R}^n$

Unless otherwise stated, "Yes" indicates a proof for the entry exists: "No" indicates that counterexamples have been found (though it may hold in some cases).

For more details of the models, see equations (4.8) - (4.13).

Condition	Trans- formation	Distance		
		$D_{1\beta} (\beta \neq 2)$	D_{12}	$D_{2\beta}$
General T.I.	-	No	Yes (*)	No
T.I. on straight lines	-	Yes, if $\beta > 1$; No, if $\beta < 1$	Yes	No
T.I. on straight lines through origin	-	Yes	Yes	No
General S.T.I.	F_1	No	No	No
	F_2	No	No	No
S.T.I. on straight lines	F_1	No	No ^(§)	No
	F_2	No	Yes ⁽⁺⁾	No
S.T.I. on straight lines through origin	F_1	No ^(§)	No ^(§)	No
	F_2	Yes ⁽⁺⁾	Yes ⁽⁺⁾	No

Footnotes (*) I have yet to find either a proof or counterexample for this.

(+) The Similarity Triangle Equality (eqtn. (4.16)) holds here: if y is between x and z on the line, then $\beta(x,z) = \beta(x,y)\beta(y,z)$.

(§) It is possible to prove that the S.T.I. holds here only trivially (see Appendix A).

The "ultrametric inequality",

$$D(x,y) \leq \max \{D(x,z), D(y,z)\} \quad \forall x,y,z \quad (4.17)$$

mentioned in Chapter 1 as a possible alternative to the Triangle Inequality cannot apply here:

consider any three real numbers α, β, γ . Suppose (without any loss of generality) that γ is the biggest of them: i.e. $\gamma \geq \max(\alpha, \beta)$. Then if the ultrametric inequality applies to these numbers for all possible arrangements of them (i.e. $\alpha \leq \max(\beta, \gamma)$; $\beta \leq \max(\alpha, \gamma)$; $\gamma \leq \max(\alpha, \beta)$) we must have $\gamma = \max(\alpha, \beta)$ so that $\gamma = \alpha$ or $\gamma = \beta$.

This shows that when (4.17) holds, the distance function D must have the very distinctive property that for any triple of points, at least two of the three interpoint distances must be equal. This is clearly not true for $D_{1\beta}$, $D_{2\beta}$, or M_{β} .

Finally, as to the criteria of Beals, Krantz and Tversky mentioned in Chapter 1, it is clear that, because of the way they are normalised, both $D_{1\beta}$ and $D_{2\beta}$ are neither interdimensionally additive, intradimensionally subtractive nor decomposable. As noted previously, they do not have segmental additivity either. This suggests that some of the axioms used by these authors are not necessarily reasonable ones.

4.1.2 Other Properties

4.1.2.1 Relationship between $D_{1\beta}$ and $D_{2\beta}$

By the Minkowski Inequality, for $\beta \geq 1$,

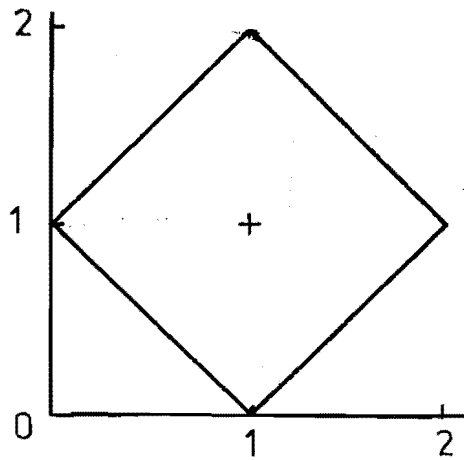
$$\left(\sum_{i=1}^n |x_i + y_i|^\beta \right)^{1/\beta} \leq \left(\sum_{i=1}^n |x_i|^\beta \right)^{1/\beta} + \left(\sum_{i=1}^n |y_i|^\beta \right)^{1/\beta}$$

so, by definition of $D_{1\beta}$ and $D_{2\beta}$,

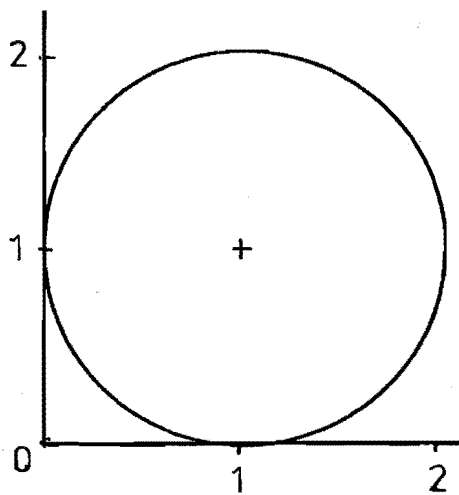
$$D_{1\beta}(x, y) \leq D_{2\beta}(x, y) \quad \forall x, y \in \mathbb{R}^n, \quad \beta \geq 1.$$

4.1.2.2 Isosimilarity contours

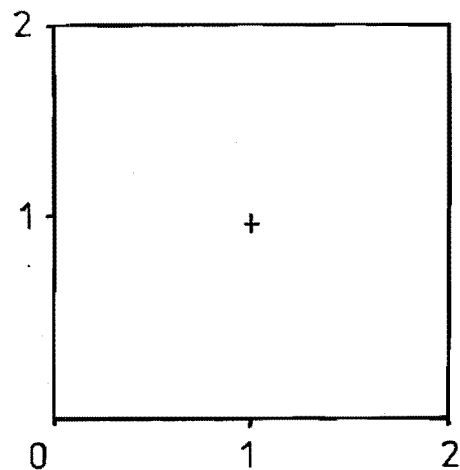
Locii of equal similarity from a given point (isosimilarity contours) have been used extensively, both in mathematics and psychology, both directly and indirectly. For the Minkowski distances, the contours are well known: they vary continuously, as β goes from 1 to ∞ , from an n -dimensional diamond, to a hypersphere (for $\beta = 2$), to an n -dimensional cube. The intermediate contours take the form of "flattened" hyperspheres or "rounded" hypercubes. The two-dimensional case is given in Figure 4.1, for $\beta = 1, 2, \infty$.



(a) 'City-Block' metric: $\beta = 1$



(b) 'Euclidean' metric:
 $\beta = 2$



(c) 'Dominance' metric:
 $\beta = \infty$

FIGURE 4.1 Isosimilarity contours for some Minkowski distances, M_β . All are centred at (1,1) and have radius = 1. See text for further details.

The contours for the Minkowski distances can be plotted without reference either to centre or radius of the contour: any contour is "typical" for a given β . This is because the translational property, D5, means that all contours of equal radius are of identical size and shape in the space, and the expansion property, D6, means that the contour of radius α about a given point is simply an expansion by a factor of α of the contour of radius 1.0 about that point.

Neither of these properties are true for $D_{1\beta}$ and $D_{2\beta}$. Except in the special case when $\beta = 2$ (see section 4.1.3), the form of the contour depends on the relative values of the coordinates of its centre. In almost all cases, the shape changes with radius. Figure 4.2 gives some representative contours for $\beta = 1$ with different radii and centres, for $\beta = 2$ with different radii, and an example for $\beta = \infty$. The contours for $\beta = \infty$ are rather complicated, being very irregular polygons often with many sides, the exact number varying with the centre and the radius. The full range of them is therefore not given.

Two obvious features of the contours in Figure 4.2 are that they are not wholly symmetric about their centres (they tend to "bulge" away from the origin), and they are not always convex, particularly

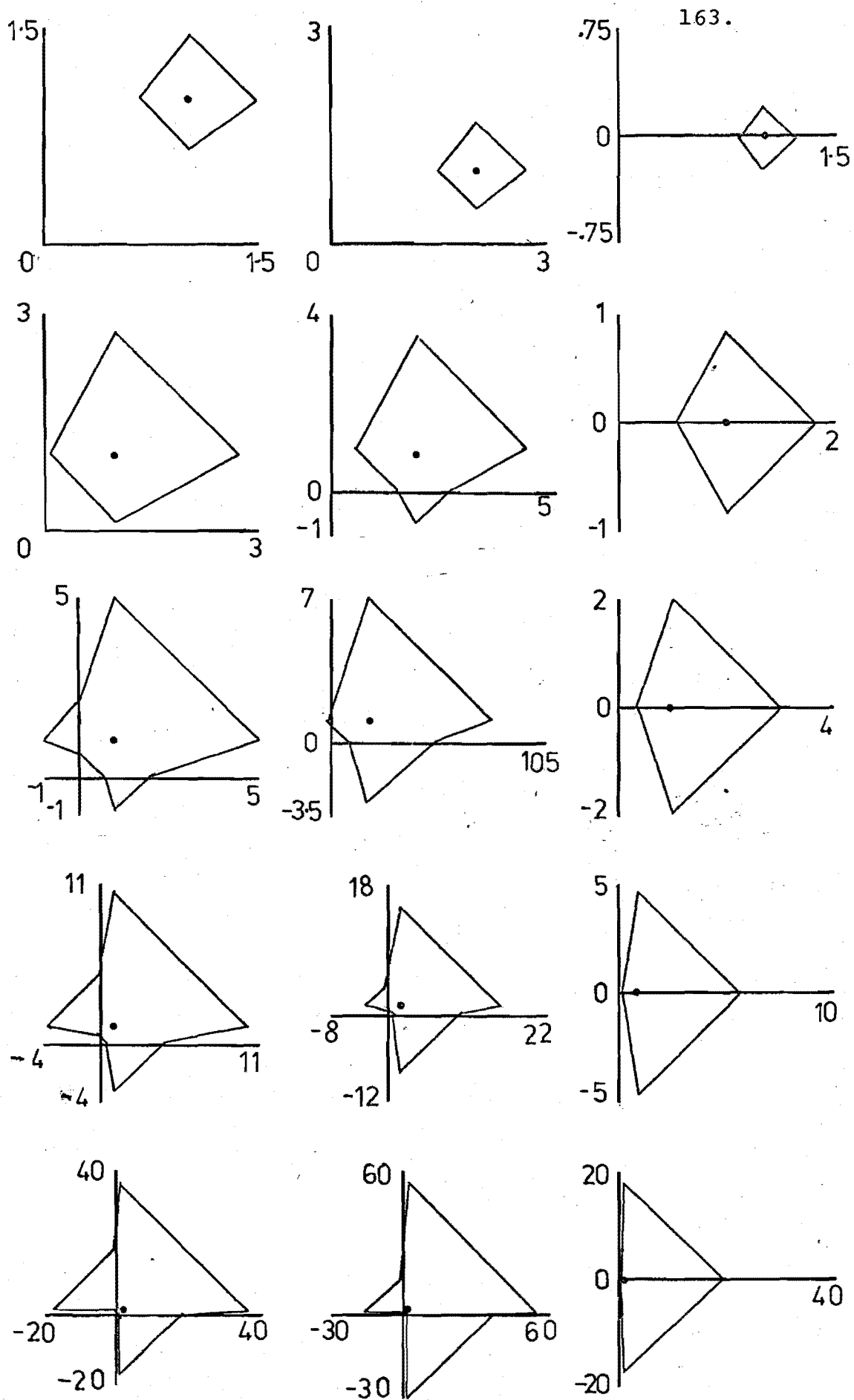


FIGURE 4.2: (a) Isosimilarity contours for D_{11} . The five rows give contours of radius $.1, .3, .5, .7, .9$ resp.; the three columns give contours centred on $(1,1), (2,1), (1,0)$, resp. Note the scales on the axes.

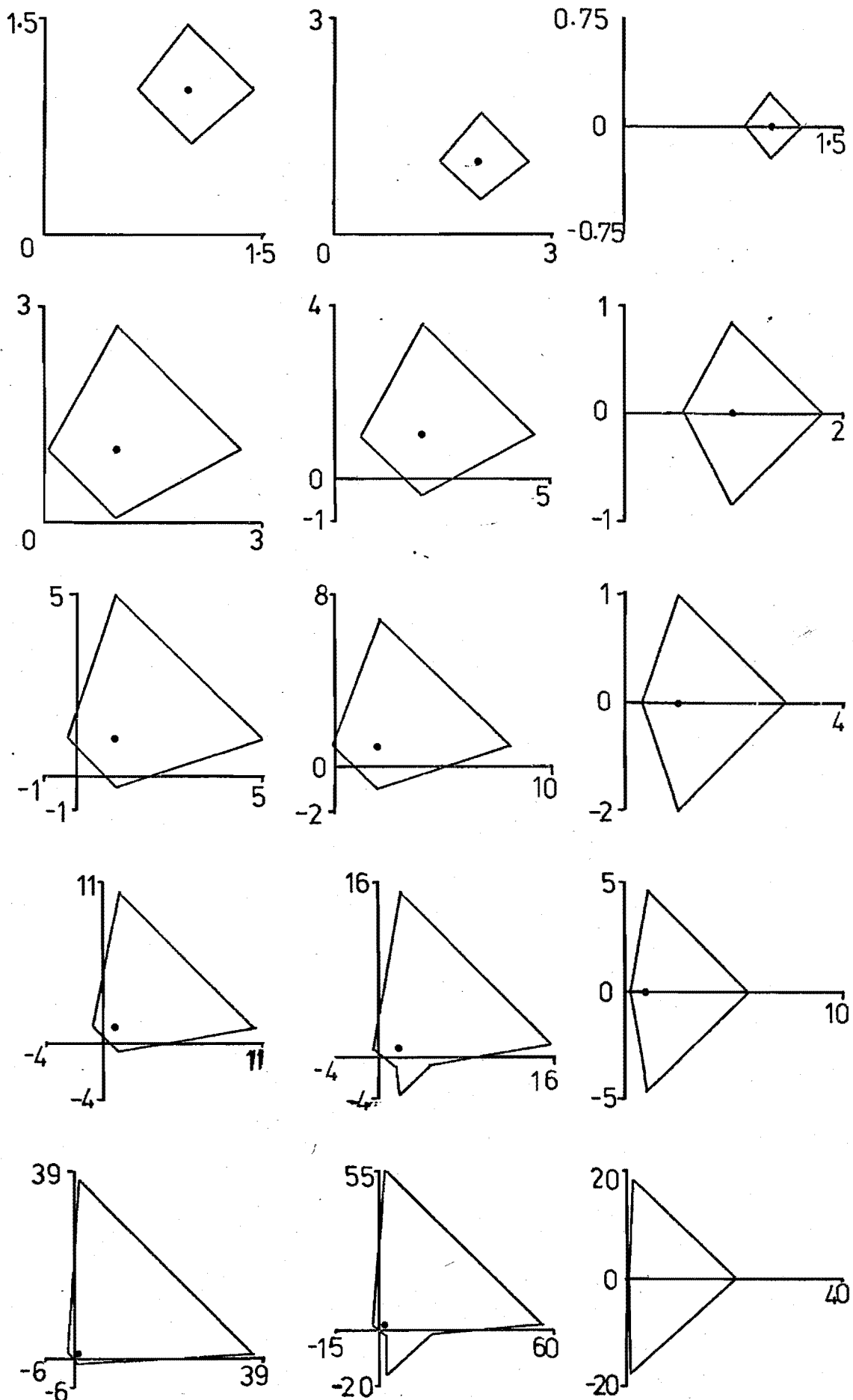


FIGURE 4.2: (b) Isosimilarity contours for D_{21} . The five rows give contours of radius .1, .3, .5, .7, .9, resp.; the three columns give contours centred on (1,1), (2,1), (1,0), resp. Note the scales on the axes.

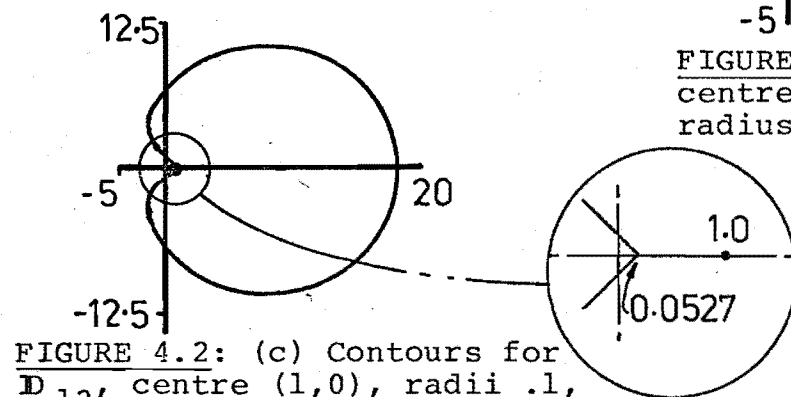
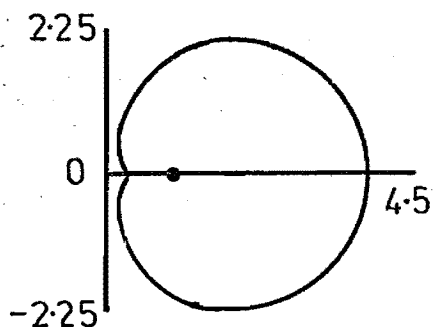
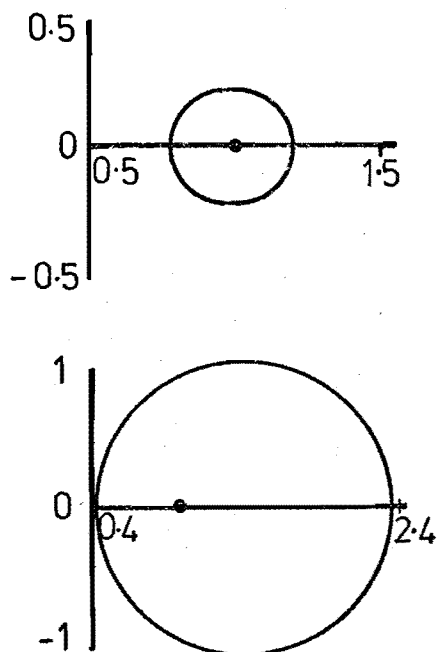


FIGURE 4.2: (c) Contours for D_{12} , centre $(1,0)$, radii $.1, .4, .6, .9$, resp.

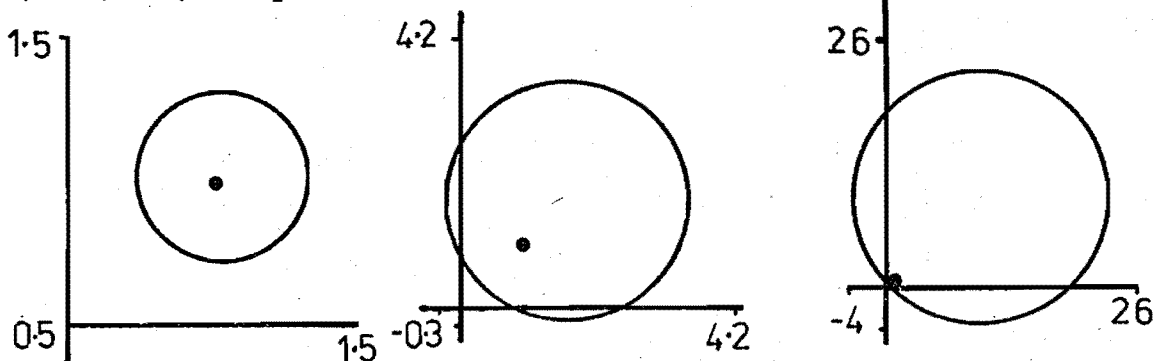


FIGURE 4.2: (d) Isosimilarity contours for D_{22} , centre $(1,1)$, radii $.1, .5, .9$ resp.

Again - note the scales on the axes.

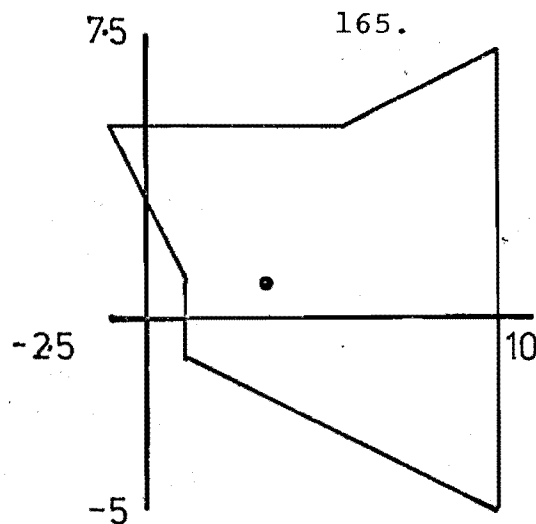


FIGURE 4.2: (e) Contour, centre $(3,1)$, for $D_{1\infty}$, radius 0.5 .

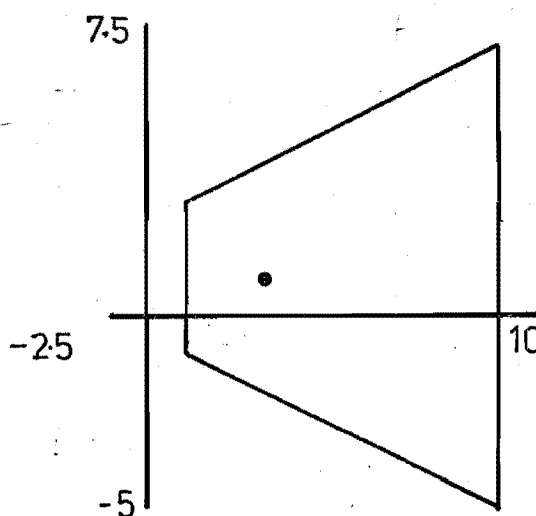


FIGURE 4.2: (f) Contour, centre $(3,1)$ for $D_{2\infty}$, radius 0.5 .

for $D_{1\beta}$ when the radius is near 1.0. These features, along with their lack of translational invariance, mean that traditional proofs of distance properties do not hold (e.g. Eggleston, 1958, p. 54ff; Beckenbach & Bellman, 1961, Chapter 6).

One point about the shapes of the contours for the normalised distances in comparison with those for the corresponding Minkowski distances is in general immediately striking: they are quite recognisably fairly simple distortions of each other. The distortion decreases (see section 4.1.2.3) as the radius gets smaller. This could explain some empirical results where Minkowski models were found to have about as good a fit to data as content models (e.g. Waern, 1970b). It also gives support for a reinterpretation of many experiments said to give results favouring a distance model. In particular, Gregson's and Eisler's set theoretic models may be expected to give ordinal results rather similar in certain aspects to those of the City-Block model, and many vector content models may have results near those of a Euclidean model. These effects will be especially noticeable for experiments using isosimilarity contours and other (particularly ordinal conditional) indirect methods. The D_{22} contours are extreme examples of this: They are merely circles (spheres, etc) whose centres have been displaced. The displacement increases with the radius.

4.1.2.3 Local behaviour

It has been previously noted that the form of the $D_{1\beta}$ and $D_{2\beta}$ distance functions is of the form of a normalised Minkowski distance, where the normalisation depends only on the two points (stimuli) whose dissimilarity is being measured. This usually means that the closer two points are together, the nearer to equal will be their contributions to the normalisation. Extending this to a collection of points which are closely grouped, their interpoint dissimilarities will be nearly proportional to the corresponding Minkowski interpoint distances. The only real exception to this is for points clustered around the origin, when points may be in opposite quadrants so that their contributions may be mutually opposing.

Therefore in general we can say that - except for near the origin - local behaviour of the present distance functions is Minkowskian. It should be remembered though that what constitutes "local" changes around the stimulus space. "Local" points may be further apart (under the Minkowski distance) far away from the origin than near to it.

For example, the local behaviour of D_{11} and D_{21} is City-Block (M_1), and of D_{12} and D_{22} is

Euclidean (M_2). This is clearly seen in the contours of the previous section: for very small radii, the contours become increasingly like the corresponding Minkowski ones: D_{11} and D_{21} contours become like City-Block contours (symmetrical, diamond-shaped); D_{12} and D_{22} contours become like Euclidean contours (circles). The further the centres of the contours are from the origin, the larger the normalised distance contour will be when it begins to look like the corresponding distance contour.

The fact that D_{12} and D_{22} , in particular, are locally Euclidean, has important implications, as this property has been assumed by several authors in various mathematical analyses (e.g. Shepard, 1957, 1960; Indow, 1974a; Luneberg, 1950; Blank, 1953; see also Luce, 1963, pp.113-115). D_{12} and D_{22} are important if only that they are credible examples of distance functions which do not satisfy the triangle inequality or segmental additivity, but which are locally Euclidean.

4.1.3 D_{12} and D_{22}

Because of their close relationship to Euclidean distance, the two distance functions D_{12} and D_{22} are of special interest. Like Euclidean distance within the

Minkowski family, these functions have distinctive properties, some of which are retained from Euclidean distance.

4.1.3.1 Locally Euclidean

As was explained in section 4.1.2.3, D_{12} and D_{22} are locally Euclidean.

4.1.3.2 Invariant under rotations about the origin

D_{12} and D_{22} are invariant under rotations of the space about its origin. This is true because Euclidean distance is invariant under rotation about any point in the space so that the numerator of both functions is rotationally invariant, while both normalisations, being dependent on the position of the origin in the space, are rotationally invariant only about the origin.

One consequence of this property is that the shape of the contours for D_{12} and D_{22} are dependent only on the radius, and not on the relative size of the coordinates of the centre. This means that one need study only contours whose centres are on a single axis. For example, in two dimensions, a D_{12} contour of

radius α , centre $(a, 0)$ (on the x-axis) has extreme points on the x-axis of $(a\frac{1-\alpha}{1+\alpha}, 0)$ and $(a\frac{1+\alpha}{1-\alpha}, 0)$ for $\alpha < \frac{1}{2}$, and $(a\frac{1-2\alpha^2}{2(1-\alpha^2)}, 0)$ and $(a\frac{1+\alpha}{1-\alpha}, 0)$ for $\alpha \geq \frac{1}{2}$ (for $\alpha > \frac{1}{2}$, a concavity between $(a\frac{1-2\alpha^2}{2(1-\alpha^2)}, 0)$ and $(a\frac{1-\alpha}{1+\alpha}, 0)$ is present). The corresponding D_{22} contour has extreme points $(a\frac{1-\alpha}{1+\alpha}, 0)$ and $(a\frac{1+\alpha}{1-\alpha}, 0)$ (it is a circle, centre $(\frac{1+\alpha^2}{1-\alpha^2}a, 0)$ radius $\frac{2\alpha a}{(1-\alpha^2)}$). By rotating the space about the origin, the contours about any point in the (two dimensional) space can be found from these two sets of contours.

4.1.3.3 Vector representation

The previous property means that D_{12} and D_{22} are representable in terms only of the lengths of the vectors corresponding to the points, and the angles between the vectors. We have, for $x = (x_1, x_2, \dots, x_n)$, $y = (y_1, y_2, \dots, y_n)$,

$$D_{12}(x, y) = \frac{(\sum_{i=1}^n (x_i - y_i)^2)^{\frac{1}{2}}}{(\sum_{i=1}^n x_i^2)^{\frac{1}{2}} + (\sum_{i=1}^n y_i^2)^{\frac{1}{2}}}$$

Calling $(\sum_{i=1}^n x_i^2)^{\frac{1}{2}}$ and $(\sum_{i=1}^n y_i^2)^{\frac{1}{2}}$, h_x and h_y respectively,

the cosine rule gives

$$(\sum_{i=1}^n (x_i - y_i)^2)^{\frac{1}{2}} = (h_x^2 + h_y^2 - 2h_x h_y \cos \phi_{xy})^{\frac{1}{2}}$$

where ϕ_{xy} is the angle between the vectors x and y (see Table 1.1 and Figure 1.1). So

$$D_{12}(x,y) = \frac{(h_x^2 + h_y^2 - 2h_x h_y \cos \phi_{xy})^{\frac{1}{2}}}{h_x + h_y} \quad (4.18)$$

A further interesting simplification is now possible.

Let $z = x - y$ and consider ϕ_{xz} , ϕ_{yz} , and $h_z (=M_z(x,y))$
 $= (h_x^2 + h_y^2 - 2h_x h_y \cos \phi_{xy})^{\frac{1}{2}}$. Then by the cosine or sine rules it is easily shown that

$$h_x = h_y \cos \phi_{xy} + h_z \cos \phi_{xz}$$

$$\text{and } h_y = h_x \cos \phi_{xy} + h_z \cos \phi_{yz}.$$

Adding these equations gives

$$(h_x + h_y) (1 - \cos \phi_{xy}) = h_z (\cos \phi_{yz} + \cos \phi_{xz})$$

so that, for $h_x \neq 0$, $h_y \neq 0$,

$$D_{12}(x,y) = \frac{h_z}{h_x + h_y} = \frac{1 - \cos \phi_{xy}}{\cos \phi_{yz} + \cos \phi_{xz}} \quad (4.19)$$

The two results corresponding to (4.18) and (4.19)

for D_{22} are somewhat less aesthetic:

$$D_{22}(x,y) = \left(\frac{h_x^2 + h_y^2 - 2h_x h_y \cos \phi_{xy}}{h_x^2 + h_y^2 + 2h_x h_y \cos \phi_{xy}} \right)^{\frac{1}{2}} \quad (4.20)$$

and, by the sine rule,

$$D_{22}(x,y) = \left(\frac{\sin^2 \phi_{yz} + \sin^2 \phi_{xz} - 2 \sin \phi_{yz} \sin \phi_{xz} \cos \phi_{xy}}{\sin^2 \phi_{yz} + \sin^2 \phi_{xz} + 2 \sin \phi_{yz} \sin \phi_{xz} \cos \phi_{xy}} \right)^{\frac{1}{2}} \quad (4.21)$$

as long as $\sin \phi_{yz}$ and $\sin \phi_{xz}$ are not both zero.

The two results (4.19) and (4.21) show clearly that the dissimilarities $D_{12}(x,y)$ and $D_{22}(x,y)$ are independent of the absolute magnitude of the vectors of x and y . They are dependent only, in a sense, on the "shape" of the triangle formed by the two points with the origin (when that triangle exists).

For $h_x = h_y$, it is easily seen that

$$D_{12}(x,y) = \sin\left(\frac{\phi_{xy}}{2}\right) \quad (4.22)$$

$$D_{22}(x,y) = \tan\left(\frac{\phi_{xy}}{2}\right) \quad (4.23)$$

4.1.3.4 Triangle inequalities

It has already been noted (see sections 4.1.1.1 and 4.1.1.3, and Table 4.1) that D_{12} appears to satisfy the triangle inequality (this can be proved at least for points on straight lines) and so appears to be a metric, and that $F_2(D_{12})$ satisfies the similarity triangle equality - that is, the similarities are multiplicative on straight lines.

4.2 THE ONE-DIMENSIONAL CASE

Suppose we have two points x, y on a straight line through the origin in \mathbb{R}^n . Then for some $a \in \mathbb{R}^n$ such that $a_1 = 1$, we have, for all $i \in \{1, 2, \dots, n\}$,

$$\begin{aligned} x_i &= a_i x_1 \\ \text{and} \quad y_i &= a_i y_1 \end{aligned}$$

Thus

$$\begin{aligned} D_{1\beta}(x, y) &= \frac{\left(\sum_{i=1}^n |x_i - y_i|^\beta \right)^{1/\beta}}{\left(\sum_{i=1}^n |x_i|^\beta \right)^{1/\beta} + \left(\sum_{i=1}^n |y_i|^\beta \right)^{1/\beta}} \\ &= \frac{\left(\sum_{i=1}^n |a_i|^\beta |x_1 - y_1|^\beta \right)^{1/\beta}}{\left(\sum_{i=1}^n |a_i|^\beta |x_1|^\beta \right)^{1/\beta} + \left(\sum_{i=1}^n |a_i|^\beta |y_1|^\beta \right)^{1/\beta}} \\ &= \frac{|x_1 - y_1|}{|x_1| + |y_1|} \end{aligned} \quad (4.24)$$

and, similarly,

$$D_{2\beta}(x, y) = \frac{|x_1 - y_1|}{|x_1 + y_1|} \quad (4.25)$$

It is immediately obvious that (4.24) and (4.25) are, respectively, just the one-dimensional forms of $D_{1\beta}$ and $D_{2\beta}$:

$$D_{1\beta}(x,y) = \frac{|x-y|}{|x|+|y|}; \quad D_{2\beta}(x,y) = \frac{|x-y|}{|x+y|} \quad \forall x,y \in \mathbb{R}.$$

Thus, on any line through the origin our two models reduce to their one dimensional forms, both of which are independent of β .

The one-dimensional cases therefore have some wider significance. They will also prove important in Chapter 5 when some problems in (one-dimensional) psychophysics are considered. In addition, the one-dimensional cases have several interesting properties which have implications spreading to the general case.

4.2.1 Incompatibility with Minkowski distances

The first use we shall make of the one-dimensional case is to give a further, more dramatic, illustration of the incompatibility - within a monotone transformation - between the Minkowski and normalised distance functions. A proof has already been given in Chapter 1 (section 1.6) but the following example reinforces that proof.

We shall restrict ourselves to \mathbb{R}^+ where, for $x \leq y$, both normalised distance functions are equal

to $(x-y)/(x+y)$. This is just an increasing function (F_2 in fact) of the ratio x/y , where $x \leq y$: i.e.

$$D(x, y) = F_2\left(\frac{x}{y}\right), \quad x \leq y \quad (4.26)^*$$

Now suppose we have a Minkowski distance function M which is a monotone increasing function of D :

$$M(x, y) = G\left(F_2\left(\frac{x}{y}\right)\right), \quad G \text{ increasing} \quad (4.27)$$

In one dimension

$$M(x, y) = H(y-x), \quad H \text{ strictly increasing}, \quad (4.28)$$

$$x \leq y$$

so, substituting from equations (4.27) and (4.26),

$$H(y-x) = F\left(\frac{x}{y}\right) \quad \forall x \leq y \quad (4.29)$$

where $F = G(F_2)$ and so is increasing. Since this is true for all $0 \leq x \leq y$, the most obvious counter-example appears when we put $x = 0$; then

$$H(y) = F(0) \quad \forall y \in \mathbb{R}^+:$$

that is, H is a constant. This contradicts the definitions of H in equation (4.28). One might object that zero values do not often occur in an empirical setting, but this does not matter since

*Note that this also implies that $D\left(\frac{1}{x}, \frac{1}{y}\right) = D(x, y)$, which has importance in scaling theory (see Torgerson, 1960; Eisler, 1962).

a parallel counterexample can be constructed for any value of x . So there is no monotone transformation that can relate the normalised distance functions D to the Minkowski distance functions M , or, more generally, to any intradimensionally subtractive distance function.

The implication to be drawn from this is that if the present models do represent dissimilarity or (monotonically transformed) similarity judgments, non-metric multidimensional scaling using a Minkowski or other subtractive model will give invalid results.

4.2.2 The Triangle Inequalities

It is proved in Appendix A (see section 4.1.1.3) that $F_2(D_{12})$ obeys the similarity triangle equality on straight lines. If one considers this property on lines through the origin, the argument given in the introduction to section 4.2 shows that in one dimension $F_2(D_{1\beta})$ also obeys the similarity triangle equality: that is if $\beta = F_2(D_{1\beta})$,

$$\beta(x, z) = \beta(x, y)\beta(y, z) \quad x \geq y \geq z \quad (4.30)$$

or, equivalently,

$$D_{1\beta}(x, z) = D_{1\beta}(x, y) + D_{1\beta}(y, z) - D_{1\beta}(x, z) D_{1\beta}(x, y) D_{1\beta}(y, z) \quad (4.31)$$

Equation (4.31) shows that, further, $D_{1\beta}$ in one dimension not only obeys the triangle inequality but also never obeys the triangle equality - is never additive - except trivially, since for $x \neq z, x \neq y, y \neq z, D_{1\beta}(x,z) D_{1\beta}(x,y) D_{1\beta}(y,z) > 0$.

The same is not true of $D_{2\beta}$ unless its domain is restricted to either only non-negative or only non-positive values. In these two cases, the two models are of course the same. Otherwise, violations of both triangle inequalities occur (under $D_{1\beta}$, the dissimilarity between points on opposite sides of the origin is always 1.0 - this guarantees the triangle inequality).

It is worth noting here that the commonly assumed condition of the triangle equality (additivity of distances) on straight lines is an extremely restrictive condition if assumed simultaneously with the condition of similarity triangle equality (multiplicativity of similarities) on straight lines, which is a temptation. It can be shown that such an assumption must lead to the model

$$\beta = e^{-\alpha D} \quad \text{for some } \alpha > 0$$

and, if in addition D is assumed to be intradimensionally subtractive,

$$\beta(x,y) = e^{-\gamma|x-y|} \quad \text{for some } \gamma > 0.$$

This latter model is not in accord with several of the important conditions for similarities given in Chapter 1. This suggests that, for prothetic continua, the additivity, multiplicativity and subtractivity conditions should be treated with caution. The former model also shows that unless one accepts the exponential form of the similarity gradient, if the similarity triangle equality on straight lines is assumed (as it is, for example, in the models due to Gregson, and to Künnapas and Künnapas, quoted in Chapter 2) then additivity on straight lines will never hold.

4.2.3 Unidimensional similarity

To look at the models of unidimensional similarity corresponding to our distance models, we shall again restrict our domain of stimuli or points to being non-negative, since it is in this domain that such models have always been defined. In this domain, the two dissimilarity models are identical. This is not to restrict the possible applicability of the present models to the whole of \mathbb{R} .

The two unidimensional models are, for $x, y \in \mathbb{R}^+$,

$$\beta_G(x,y) = F_2(D(x,y)) = \frac{1 - \frac{|x-y|}{x+y}}{1 + \frac{|x-y|}{x+y}}$$

$$\beta_E(x,y) = F_1(D(x,y)) = 1 - \frac{|x-y|}{x+y}$$

so,

$$\beta_G(x,y) = \frac{x+y-|x-y|}{x+y+|x-y|} = \frac{\min(x,y)}{\max(x,y)} \quad (4.32)$$

$$\beta_E(x,y) = \frac{x+y-|x-y|}{x+y} = \frac{2\min(x,y)}{x+y} \quad (4.33)$$

Equation (4.33) is the Eisler-Ekman unidimensional similarity equation. The alternative unidimensional similarity equation suggested by various authors (see sections 2.1.1.1, 2.1.3.3 and 2.1.3.4) is that β is a power function of the ratio between physical stimulus values. Equation (4.32) states that β_G is simply the ratio between subjective estimates of the stimulus values. Thus, if the subjective estimates in question obey Stevens' Power Law, then (4.32) states that β_G is a power of the (physical) stimulus ratio, with exponent equal to Stevens' exponent for that particular continuum. Data analysed by Künnapas and Künnapas (1973) does not support this value of the exponent. Thus (4.32) needs modifying by some power transformation (e.g. Künnapas and Künnapas, op. cit.); with or without modification, though, it on the whole retains the same basic mathematical properties and is worth investigation.

4.3 SUMMARY

It seems useful to summarise here the more important "gross features" of the distance functions $D_{1\beta}$ and $D_{2\beta}$ that we have found, and to pull together the strings of the chapter in order to clarify the perspective and intuitions which it set out to obtain.

For $D_{1\beta}$ the most important features corresponding to those in Chapter 1 are:

$$0 \leq D_{1\beta}(x,y) \leq 1 \quad \forall x,y \in \mathbb{R}^n \quad (4.34)$$

(cf. S1)

$$D_{1\beta}(x+a,y+a) < D_{1\beta}(x,y) \text{ if } x,y,a \text{ are in the same "quadrant" of } \mathbb{R}^n, \quad a \neq 0 \quad (4.35)$$

(cf. S5)

$$D_{1\beta}(\alpha x, \alpha y) = D_{1\beta}(x,y) \quad \forall x,y \in \mathbb{R}^n, \alpha \in \mathbb{R} \quad (4.36)$$

(cf. S6)

and for $D_{2\beta}$:

$$\begin{aligned} 0 \leq D_{2\beta} \leq 1 & \quad \forall x,y \in \mathbb{R}^{n+} \\ 0 \leq D_{2\beta} < \infty & \quad \forall x,y \in \mathbb{R}^n, x \neq -y \end{aligned} \quad (4.37)$$

(cf. S1 and D1)

$$D_{2\beta}(x+a,y+a) < D_{2\beta}(x,y) \text{ if } x,y,a \text{ are in the same "quadrant" of } \mathbb{R}^n, \quad a \neq 0 \quad (4.38)$$

(cf. S5)

$$D_{2\beta}(\alpha x, \alpha y) = D_{2\beta}(x, y) \quad \forall x, y \in \mathbb{R}^n, \alpha \in \mathbb{R} \quad (4.39)$$

(c.f. S6)

These properties, along with others discussed in the chapter, lead to the following general description of the two distance functions.

$D_{1\beta}$ and $D_{2\beta}$ are normalised distances or normalised distance functions. This means that in certain important portions of the space, in the case of $D_{2\beta}$, and in the whole space, in the case of $D_{1\beta}$, they are bounded above by 1.0. It also gives two important characteristics to the functions. Firstly, pairs of points which differ by the same vector have a decreasing dissimilarity as they get further from the origin of the space (see (4.35) and (4.38)). A particular example of this is that grid lines in Cartesian coordinate space will appear to be closer together as they get further from the origin. Loosely: distant pairs of points are less dissimilar than close pairs of points.

Secondly, linear expansions of the space about its origin make no difference to interpoint dissimilarities. This is in distinction to the distance property (D6) which increases dissimilarities in proportion to the expansion. Loosely: pairs of points in equal proportion will be equally dissimilar.

Shepard (1960) suggested that psychological space might turn out to be conceived of most simply as an embedding in a Euclidean space, like, for example, a sphere. In two or three dimensions, the present distance functions are very reminiscent - though not completely accurate representations - of the distance function imposed on two- or three-dimensional Euclidean space by perspective in monocular vision. This may at least be a fruitful analogy from which to work.

$D_{1\beta}$ and $D_{2\beta}$ fail to be distances through properties (4.35) and (4.36), and (4.38) and (4.39), respectively. However, they fail additionally, with the probable exception of D_{12} , in that they do not in general satisfy the triangle inequality. In many cases, though, especially for $D_{1\beta}$ and $\beta \geq 1.5$, the violations of the inequality are small enough for it to be a reasonable approximation and insight into the way these functions behave. For $D_{1\beta}$ the triangle inequality is obeyed on straight lines.

A stronger and more important point about their distance properties, as far as non-metric MDS is concerned, is that neither distance function can be expressed as a monotone function of any distance: and so, in particular, Minkowski distance.

Finally, $D_{1\beta}$ and $D_{2\beta}$ impose a particular structure on the axes and origin of the space. Every point in the space has a dissimilarity of 1.0 from the origin: all points are "totally different" from the origin. Dissimilarities between points on different axes are also 1.0 in the case of $D_{2\beta}$, and near to 1.0 in the case of $D_{1\beta}$: loosely, axes are totally, or very, dissimilar. This could be interpreted as independence of dimensions.

Turning briefly to the similarity models, with the two similarity gradients

$$F_2(x) = \frac{1-x}{1+x} \quad 0 \leq x \leq 1$$

$$\text{and } F_1(x) = 1-x \quad 0 \leq x \leq 1,$$

both models reduce to, respectively, the Gregson and Eisler-Ekman similarity models when points are restricted to the positive "quadrant". $F_2(D_{12})$ on straight lines, and the one-dimensional case both obey the similarities triangle inequality,

$$\beta(x,y) \geq \beta(x,z)\beta(z,x) \quad \forall x,y,z.$$

In other respects the similarity functions $F_1(D_{1\beta})$, $F_2(D_{1\beta})$, $F_1(D_{2\beta})$ and $F_2(D_{2\beta})$ have properties parallel with $D_{1\beta}$ or $D_{2\beta}$.

On the whole, $D_{1\beta}$ is - at least mathematically - the more attractive of the two functions: it is bounded and well-defined (except at the origin) and more closely obeys the triangle inequality than $D_{2\beta}$. It also has the pleasant property that "opposite points are opposite": points on opposite sides of the origin, and colinear with it, are totally dissimilar - have a dissimilarity of 1.0 - under $D_{1\beta}$.

CHAPTER 5SOME PROBLEMS OF PSYCHOPHYSICS FROM A SIMILARITYVIEWPOINT

In this chapter, the models of Chapter 3 will be applied to a few of the more theoretical problems of psychological measurement not directly related to the psychology of similarity itself. Most of them have not arisen from a purely theoretical context, but from the interaction between various models and the data the models are supposed to describe or explain.

The data are mainly drawn from unidimensional psychophysics and we shall be dealing with the problems of category scales almost exclusively. Two other areas will be covered briefly at the end of the chapter, more to exhibit the uses and weaknesses of present and other models than to make a substantial contribution towards solving the particular problems.

It will be recalled from Chapter 3 that we have three basic dissimilarity models:

$$D_{1\beta}(x,y) = \frac{(\sum |x_i - y_i|^\beta)^{1/\beta}}{(\sum |x_i|^\beta)^{1/\beta} + (\sum |y_i|^\beta)^{1/\beta}} \quad \forall x,y \in \mathbb{R}^n \quad (5.1)$$

$$\text{and } D_{2\beta}(x,y) = \frac{\sum |x_i - y_i|^\beta}{\sum |x_i + y_i|^\beta} \quad \forall x,y \in \mathbb{R}^n \quad (5.2)$$

for prothetic continua, and

$$M_\beta(x,y) = (\sum n_i |x_i - y_i|^\beta)^{1/\beta} \quad \forall x,y \in \mathbb{R}^n, \quad (5.3)$$

$$|x_i - y_i| \leq 1, \quad i=1, \dots, n;$$

$$n \in \mathbb{R}^{n+}, \quad \sum n_i = 1$$

for equal-measure metathetic continua, where $\beta > 0$ in each case. Note that $M_\beta(x,y) \leq 1$.

We also have two "similarity gradients" which may transform the dissimilarity models into similarity models:

$$F_1(x) = 1-x \quad 0 \leq x \leq 1 \quad (5.4)$$

$$\text{and } F_2(x) = \frac{1-x}{1+x} \quad 0 \leq x \leq 1 \quad (5.5)$$

F_1 and F_2 are strictly decreasing (and hence one-to-one).

In this chapter, models (5.1), (5.2) and (5.3) will be used mainly in their one dimensional cases; for these we shall use the notations

$$D_1(x,y) = \frac{|x-y|}{|x|+|y|} \quad (5.6)$$

$$D_2(x,y) = \frac{|x-y|}{|x+y|} \quad (5.7)$$

$$M(x,y) = |x-y| \leq 1 \quad (5.8)$$

respectively. For $x, y \in \mathbb{R}^+$, $D_1 = D_2$; we shall in that case sometimes put $D = D_1 = D_2$.

5.1 BISECTION, AND CATEGORY SCALES

5.1.1 Bisection.

In the bisection experiment, two stimuli (generally unidimensional: say x and z) are presented to the subject who is asked in some way to indicate a third stimulus (y) which is "halfway between" or "produces equal intervals with" the first two (e.g. Stevens, 1960a, p.39).

The notion "halfway", if taken literally, makes sense only if the distance function is additive since it implies that we have, for a distance function d ,

$$d(x,y) = d(y,z) = \frac{1}{2}d(x,z)$$

so that

$$d(x,z) = d(x,y) + d(y,z).$$

This (as has been pointed out in Chapter 4) is not in general true for D_1 and D_2 , although a partial exception will be discussed below. For example, if $x, z \in \mathbb{R}^+$, $x < z$, then for $y \in \mathbb{R}^+$ to be such that

$$D(x, y) = D(y, z) = \frac{1}{2}D(x, z)$$

we must have

$$\frac{y-x}{y+x} = \frac{z-y}{z+y} = \frac{z-x}{2(z+x)}$$

which can only happen if $x = y = z$, a trivial case.

We must therefore make use of the "equal interval" notion (which coincides with the "halfway" interpretation if the distance function is additive) by taking the bisection point, \overline{xz} , of x and z to be defined by

$$d(x, \overline{xz}) = d(\overline{xz}, z).$$

For $x, z \in \mathbb{R}^+$, or $-x, -z \in \mathbb{R}^+$, and $x, z \neq 0$, $z > x$, this gives, for D_1 and D_2 ,

$$\overline{xz} = \sqrt{xz} \quad (5.9)$$

which is clearly unique. With the restrictions on

x and z given above, this gives a consistent bisection operation (see, for example, Restle, 1961, p. 197-199; Marks, 1974, p. 249ff; Pfanzagl, 1968, p. 84).

Problems occur though when we have $z > 0$, $x < 0$, or when $x = 0$ or $z = 0$. If the latter occurs - say $x = 0$ - then

$$\forall y \in \mathbb{R}, \quad D_1(x, y) = D_2(x, y) = 1$$

so clearly no bisection can take place since we have defined $D_1(0, 0) = D_2(0, 0) = 0$. We must therefore exclude the origin from consideration as a standard stimulus in the bisection operation.

The other problem occurs when x and z are on opposite sides of the origin: $x < 0$, $z > 0$. It can then be shown that, uniquely,

$$\overline{xz} = 0 \quad (5.10)$$

when $D_1(x, \overline{xz}) = D_2(x, \overline{xz}) = D_2(\overline{xz}, z) = D_1(\overline{xz}, z) = 1$.

This means that a consistency test cannot be carried out, since we have excluded the origin as a candidate for either of the two stimuli to be bisected. This does not mean that the bisection definition is inconsistent: it means only that consistency does not have a real meaning.

To avoid these problems, we shall therefore consider only points in \mathbb{R}^+ in this section (section 5.1). We still have the difficulty (if it is a difficulty) that any interval with the origin as endpoint cannot be bisected; this appears to be an integral part of the model \mathcal{D} which should be taken as a prediction of a certain behaviour. This section, then, deals only with positive, unidimensional stimuli, unless otherwise stated.

No such difficulties appear, in the metathetic case, with \mathcal{M} . The unique bisection point there is of course

$$\overline{xz} = (x+z)/2 \quad (5.11)$$

no matter where x and z are in \mathbb{R} (as long as $|x-z| \leq 1$).

An interesting point, using the above definition of bisection, which may relate the two models \mathcal{D} and \mathcal{M} , is a consequence of the normalised nature of the two models $\mathcal{D}_{1\beta}$ and $\mathcal{D}_{2\beta}$.

When we talk about the one-dimensional case, we generally take it to mean one dimension on one of the axes of a multidimensional space, so that other dimensions have zero values. But with the normalised

distance functions, it makes a difference to talk about one dimensional cases not on an axis. One such case that immediately comes to mind is the line parallel to an axis; in two dimensions

$$L_a = \{(x, a) : x \in \mathbb{R}\} \quad \text{for some } a \in \mathbb{R} \quad (5.12)$$

In this case we have that if $x_a, z_a \in L_a$,
 $x_a = (x, a) \quad z_a = (z, a), \quad z > x$, then

$$D_{1\beta}(x_a, z_a) = \frac{z - x}{(|x|^\beta + |a|^\beta)^{1/\beta} + (|z|^\beta + |a|^\beta)^{1/\beta}} \quad (5.13)$$

$$D_{2\beta}(x_a, z_a) = \frac{z - x}{(|x+z|^\beta + 2^\beta |a|^\beta)^{1/\beta}} \quad (5.14)$$

although of course, for any β ,

$$M_\beta(x_a, z_a) = z - x.$$

Suppose now we wish to find the bisector,

$\overline{x_a z_a} = (\overline{xz}, a)$ of the points x_a, z_a on L_a : then for $D_{1\beta}$ we have, from (5.13),

$$\begin{aligned} & \frac{z - \overline{xz}}{(|\overline{xz}|^\beta + |a|^\beta)^{1/\beta} + (|z|^\beta + |a|^\beta)^{1/\beta}} \\ &= \frac{\overline{xz} - x}{(|\overline{xz}|^\beta + |a|^\beta)^{1/\beta} + (|x|^\beta + |a|^\beta)^{1/\beta}} \quad (5.15) \end{aligned}$$

The solution to this is not easily obtained; but suppose we let $|a|$ get very large; then the denominators of both sides of (5.15) get closer and closer to being equal to $2|a|$, so we can cancel giving the limiting case of

$$z - \overline{xz} = \overline{xz} - x$$

or

$$\overline{xz} = \frac{x+z}{2} \quad (5.16)$$

An identical result follows from (5.14).

This is of course the same result we found for M (equation 5.11): in other words

$$\lim_{|a| \rightarrow \infty} (\text{bisector under } D_{1\beta} \text{ or } D_{2\beta}) = \text{bisector under } M \quad (5.17)$$

What then does "a" represent? Looking back to our original construction in Chapter 3, "a" is related to the "core" of elements in common to all stimuli. What (5.17) says is that as this core becomes more and more dominant - perhaps as the core becomes more and more "obvious" - prothetic dissimilarity becomes mathematically more and more like metathetic similarity. In the limit it becomes additive, or

in a sense which will become clear in the next section, linear. Empirically this could lead to confusion between the prothetic and metathetic cases.

It will be recalled from Chapter 2 that equation (5.9) for the bisection point has been rejected by various authors (e.g. Luce, Stevens and Galanter) because (a) it is inaccurate and (b) it does not take account of bias and hysteresis effects. Along with Luce (1961) one could include additional parameters to take these objections into account, but one would like to avoid that if possible, since it seems rather ad hoc. A better answer may be through the previous discussion of "core": the subject's perception of the size of the core may vary with order of presentation or with payoffs.

However, a third answer may come from an observation of Restle's (Restle, 1961, p.216) on category scales for prothetic continua, that "if the subject is given enough categories to prevent artificial bunching, if he judges a dimension which can be estimated quite accurately....., and if the stimuli are spaced equally in physical amount to minimise distortions from context, one finding is consistent - the middle category is used for a stimulus which is one-third of the way from the smallest to the largest stimulus".

To Restle's experimental preconditions, we might add that of uniform distribution of stimuli presented (see Parducci, 1965); but if we take the middle category to be approximately the same as the position of a bisection judgment, as seems plausible (see Restle, op.cit., but for contrary evidence, see Adams and Fagot, 1975), this observation leads directly to the hypothesis (for prothetic continua) that

$$\overline{xz} = x + \frac{z-x}{3} = \frac{2x+z}{3} \quad \forall x, z \in \mathbb{R}^+, z > x \quad (5.18)$$

This form of bisection equation is basically quite different from equation (5.9) and needs a new approach to the interpretation of what the bisection judgment consists of. Note that it is not a consistent definition in that

$$\overline{(x \overline{xz})} \quad \overline{(\overline{xz} z)} \neq \overline{xz},$$

but this is in accord with results quoted by Marks (op.cit).

The value represented by equation (5.18) is dependent on two other values: one of the endpoint stimuli, and the absolute difference between them. This suggests that the difference between the endpoints is an integral part of the bisection judgment. A possible way it could be incorporated into a

judgment process is as follows: the difference itself is to be regarded as a stimulus; in other words, the zero point is moved from the origin to the smaller or lesser of the two stimuli to be bisected. The bisection judgment is then to find a stimulus with a dissimilarity of one-half to this "difference" stimulus. To restore relativities, when the bisector is found, the origin is restored to its original position.

Mathematically, this process is represented by

$$D(\overline{xz} - x, z - x) = \frac{1}{2} \quad \forall x, z \in \mathbb{R}^+, z > x \quad (5.19)$$

Solving this for \overline{xz} gives

$$\frac{(z - x) - (\overline{xz} - x)}{(z - x) + (\overline{xz} - x)} = \frac{1}{2}$$

or

$$\overline{xz} = x + \frac{1}{3} (z - x) \quad (5.20),$$

the required result.

However this is not nearly as straightforward as it may appear. Can a difference be used by a subject as a stimulus as easily as the process would assume? Can a subject shift the origin in this way?

Why should the origin be shifted to the smaller stimulus: why not to the larger stimulus, with a reversal of the direction of the dimension? Why should the judgment involve dissimilarity and not similarity? The ratio (Gregson's) unidimensional similarity model gives a different result (it halves the interval). And finally, but perhaps most importantly, the dissimilarity models have up to now been assumed to be in psychological - not physical - space. The result (5.20) is thus for psychological space, whereas Restle's result (5.18) was for physical space; since the psychophysical function, whatever it is, often deviates significantly from identity, the two results may in fact be hard to reconcile.

A few preliminary answers might be given along the following lines. The "shifting of the origin" is not necessarily an actual process carried out by the subject: it is more a conceptual way of looking at the process for mathematical convenience. On the other hand, it may be a quite simple process of ignoring all "small" aspects of the stimuli. The question of why the smaller stimulus should be used is not easily answered - it may be more "natural" to subjects in some way. The dissimilarity-similarity question is not a real one if the Eisler-Ekman similarity

model $F_1(D)$ is assumed, since this gives identical results to the dissimilarity model; usually though, similarity is considered a more "natural" judgment to make than dissimilarity. Lastly, the relevance of the question of the space we are dealing with depends on (a) the form of the psychophysical function, and (b) the way (and whether) the difference judgment is made.

We shall for the present ignore these important objections for the purely pragmatic reason that the approach gives good results. The next section will provide further evidence for this attitude.

Finally it should be noted that the above analysis was for the prothetic continuum case. An exactly parallel analysis can be carried out for the metathetic case being studied here, with the result (using the same controversial assumptions as before) that the bisection judgment should halve the interval between the two stimuli.

5.1.2 Category Scales

Category scales are generally thought of as approximating to "equal interval" scales (e.g. Stevens and Galanter, 1957, p. 377) in the same sense as the bisection judgment divides an interval into two "equal intervals". The subject is instructed in some way to make category widths "psychologically equal". As explained in section 2.3.3, several workers, including Ekman, Goude and Waern, (1961) and Junge (1960), have already given this a similarity interpretation by equating "psychologically equal" with "of equal similarity". Under the present models for prothetic unidimensional continua (and any other models which are simply functions of the stimulus ratio), this means that all pairs of stimuli equally spaced on the category scale are in equal ratio. From this simple basis, one can construct the category scale without any further assumptions, and show that it is logarithmically related to the psychophysical scale on which the similarity judgments are based.

In fact, of course, subjects are rarely instructed to consider all possible intervals: rather, they are asked (if asked explicitly at all) to in some way "partition a segment of the continuum into equal intervals" (Stevens and Galanter, op.cit.). In

other words, they are asked to make equal only the intervals either (depending on one's interpretation) between adjacent category boundaries (limens), or between consecutive category "exemplars" (which we shall define as the stimuli corresponding to the respective category labels.) There could therefore be some debate over whether boundaries or exemplars should be the critical values used or solved for; in practice it would probably make very little difference to the fit of a model. Thurstone used limens (see Torgerson, 1958, for example); we shall use exemplars because they seem more natural and less abstract in the context of similarity judgments. Stevens and Galanter (op.cit. p. 392) showed exemplars could be found by subjects when asked for; this procedure is closely related to that of Equissection (e.g. Torgerson, 1958, Chapter 6); and see also Reed (1972), for evidence on use of exemplars.

In Appendix B it is proved that, for the exemplars, the category scale is logarithmically related to the stimulus values. The important point about this particular proof is that it needs no more than the equal ratios assumption for its conclusions: no extra assumptions are required (compare the proof of Junge, 1960, or Eisler, 1962).

It also makes clear the fact that the logarithmic scale is valid only for the actual category exemplars: intermediate stimuli are, by some decision rule, put by subjects into categories. These are then averaged by the experimenter to give a non-integral scale value which will depend on the decision rule used and the distribution of discriminial processes of the subject (compare the much simpler, but deceptive, proof of Wagenaar, 1975). Of course, if consecutive exemplars are relatively closely spaced, then a logarithmic relation will be a good approximation for all values.

The scale then becomes

$$K(x) = \begin{cases} 1 & x \leq a \\ (n-1) \frac{\log(\frac{x}{a})}{\log(\frac{b}{a})} + 1 & a \leq x \leq b \\ n & x \geq b \end{cases} \quad (5.21)$$

where $K(x)$ is the category scale value of stimulus x , and a, b are the exemplars of the first and last (n th) categories respectively.

The exemplars a and b will not necessarily be identical to the extremes of the stimulus sample: subjects may set up their own, perhaps according to various experimental conditions such as stimulus density, and frequency and order of presentation.

They can be estimated from the results of a category-scaling experiment. Using Stevens (1957) and Stevens and Galanter (1957) as bibliographies. of category-scaling experiments, this estimation was carried out approximately on a number of such experiments. The results are given in Table 5.1 along with an F-ratio expressing the linearity between the scale data and the scaling model with the tabulated end exemplars. It will be seen that the fit is generally good, and highly significant if the usual homoscedastic normality assumptions are true of the data. The estimated end exemplars are plausible, though deviating widely at times from the actual range of stimuli presented.

The second possible model of category scales corresponds to that leading to equation (5.20), for the bisection model. This suggests that a category scale is essentially a dissimilarity (or similarity) scale such that, using the notation of equation (5.21),

$$g(x-a, b-a) = \begin{cases} \frac{K(x)-1}{n-1} & x \leq b \\ 1 & x > b \end{cases} \quad (5.22)$$

TABLE 5.1

F-ratios and fitted parameters for two Category Scaling Models.

Data used is all referred to in Stevens (1957) and Stevens and Galanter (1957: referred to below as "SG"). For details of models see text.

Continuum; Stimulus Spacing		Stimulus Range	Log Model		Similarity-Scale Model		d.f.	No. of categ's	Data Reference
			Estimated End Exemplars	F-Ratio	Estimated End Exemplars	F-Ratio			
<u>Weight;</u>	Equal	90-400	58,498	8648	0,488	4331	3	9	Helson (1948,p.305)
	Equal	200-400	141,444	3880	146,455	12930	3	9	Helson (1948,p.301)
	Equal	200-900	155,729	17137	147,717	3449	3	9	"
	Equal	90-400	64,553	3364	8,536	1238	3	9	"
	Equal	50-100	45,108	7376	47,113	6367	19	10	Guilford and Dingman (1955)*
	Equal	50-100	33,132	7376	32,137	7332	19	15	"
	Equal	84-108	77,110	628	80,113	462	5	3	SG(Fig 5A: Fernberger's results) [†]
	Equal	50-850	82,981	530	5,904	3490	7	11	Nash (1950)
	Equal	50-850	80,823	6110	3,814	404	7	9	"
	Equal	50-850	81,833	1239	4,822	381	7	7	"
	Equal	50-850	90,999	455	38,908	704	7	5	"
	Equal	50-850	107,760	4364	68,747	1705	7	3	"
	Veg	43-292	61,325	846	42,329	772	9	11	Canter and Hirsch (1955, Fig. 1) [†]
	Equal	43-292	42,370	286	36,329	1249	9	11	"
	Log	43-292	40,295	3315	26,289	636	9	11	"
	Equal	84-108	79,112	1494	82,116	492	5	3	SG(Fig 5A: Wever & Zener data) [†]
	Equal	88-104	82,109	361	85,112	136	3	3	"
<u>Duration;</u>	"Pure"	0.25-4	5,4.8	795	3,4.4	5766	11	7	Stevens (1957, Fig 1) [†]
	Bunched down	250-1000	230,1268	366	223,1209	980	3	5	Postman and Miller (1945)
	"	250-1000	225,1370	553	207,1294	7936	3	5	"
	"	250-1150	225,1393	443	280,1307	3116	3	5	"
	"	250-1400	226,1633	207	203,1468	760	3	5	"
	"	250-1500	225,1735	304	198,1531	2668	3	5	"
	Equal	0.25-2.5	4,3.1	563	4,2.9	1454	8	7	SG(Fig 2A: their data) [†]
	Equal	0.25-4	5,4.5	560	3,4.1	2182	14	7	"
	Log	0.4-4	5,4.3	824	3,4.0	1790	9	7	SG(Fig 2B: their data) [†]
	Bunched upwards	0.25-4	7,4.9	404	6,4.6	609	8	7	"
<u>Loudness;</u>	Log	1-64	1.6,63	702	0,37	736	11	7	SG(Fig 8B: numerical scale) [†]
	Log	1-64	1.7,59	230	0,39	458	10	3	SG(Fig 8C: 3 pt. scale) [†]
	Log	7-64	1.7,93	468	0,49	490	12	100	SG(Fig 8C: 100 pt. scale) [†]
	Equal	2-46	4,49	440	1,47	1051	8	7	SG(Fig 9A: some spacing) [†]
	"Pure"	1-64	2.4,70	562	0,50	598	9	7	SG(Fig 9A: "adjusted") [†]
	S finds Exemplars	3-34	4.4,36	772	3,34	1423	5	7	SG(Fig 9B: S adjusts stimuli) [†]
	Equal	3-34	4.8,38	2068	4,35	4480	11	7	SG(Fig 9B: S assigns category) [†]
	?	4-64	8,69	1223	0,27	107	5	7	SG(Fig 10A: Reger's scale) [†]
	Bunched upwards	4-64	1.2,98	242	0,45	86	5	7	SG(Fig 10A) [†]
	Bunched down	4-64	6,79	438	0,27	186	5	7	"
<u>Length;</u>	Bunched to ends	3-112	12,118	553	7,110	2303	15	11	SG(Fig 1) +
<u>Numerousness;</u>	Log	15-74	13,88	2406	10,84	1473	21	9	Guilford (1954, p.203)
<u>Area;</u>	Bunched down	4-81	5,83	828	1,69	724	7	5	SG(Fig 3B: their data) [†]

Footnotes:

* A 15-point scale was used in this experiment, of which only 10 pts. were actually used significantly often. The first result above rescales to a 10 pt. scale; the second uses the actual data.

[†] These use data merely estimated from the referenced graph.

Our previous experience would suggest that " β " here is β_E ; in this case the model is

$$K(x) = \begin{cases} 1 & x \leq a \\ 2(n-1)\frac{x-a}{x+b-2a} + 1 & a \leq x \leq b \\ n & x \geq b \end{cases} \quad (5.23)$$

This is concave downwards, as required. We shall refer to it as the "similarity-scale model" of category scaling, to distinguish it from the (albeit similarity based) logarithmic model (5.21).

Again, exemplars a and b were estimated from the data, as shown in Table 5.1. They are, again, plausible though at times widely divergent from the respective stimulus ranges.

The fit of the similarity-scale model is also very good, the F-ratios being in general at highly significant levels. But there is very little to choose between it and the logarithmic model. There seems to be little pattern in the fits of the two models; the only immediately identifiable tendency being for the worst fits to occur when the number of available categories is very high (i.e. 100) or rather low (i.e. 3 or 5).

In view of the previous discussion of bisection, and the general discrediting of the logarithmic model (see, for example, Stevens, 1966), the similarity-scale model of category scaling seems attractive. It may perhaps be taken as indirect evidence for the Eisler-Ekman model of unidimensional similarity, over that of Gregson - but only if and when all the assumptions it requires for its construction are validated in some way. Its approach is quite similar to one of the models of Helson's Adaptation Level theory. In fact, if the category scale is taken to be a scale of dissimilarity from a single "Adaptation Level" stimulus, then the model is

$$K(x) = k \frac{|x-a|}{x+a}$$

(where a is the "Adaptation Level" and k is a suitable unit) which is formally identical to that given in Helson (1948, p 303). Although the present similarity-scale model has two interpretable parameters instead of Helson's one, the close relationship between the two models may mean that our model suffers from many of the well-known weaknesses that Helson's possesses (see, for example, Stevens, 1958; Parducci, 1965).

The most attractive feature, however, of the present similarity-related approach to category scaling is that it explains the difference between the form of the category-scale function for prothetic and metathetic continua. The curvilinear functions derived as equation (5.21) and (5.23) are true only for prothetic continua, since the similarity models used in their derivations are assumed to apply only there. The corresponding similarity model for (unidimensional) equal-measure metathetic continua is a function of the absolute difference between stimulus values (see equations 5.3 and 5.8):

$$g_E(x,y) = 1 - |x-y| \quad (5.24)$$

$$g_G(x,y) = \frac{1 - |x-y|}{1 + |x-y|} \quad (5.25)$$

If model (5.24) is the more correct, as evidence from the prothetic case suggests, the form of the category-scale function will be linear, whichever of the two scale constructions is used. This is in line with the suggestion by Stevens and Galanter who state (p. 377, op.cit.) that category scales on metathetic continua "may be linear" when plotted against the ratio scale of subjective magnitude. In terms of the present models, non-linearity could be due to the continuum not being purely metathetic - that is, not an exactly "equal-measure" continuum - but being somewhere along the continuum to protheticness.

Thus, if the line of reasoning presented in this and previous chapters is correct, an explanation for this particular aspect of the difference between metathetic and prothetic continua will have been given. We are not aware of any other explanation.

5.2 OTHER PROBLEMS

5.2.1 Helm's result

Helm and his co-workers (see, for example, Coombs, 1964, p. 488ff., and Helm, Messick and Tucker, 1961) have noted that both Stevens' and Galanter's relation between category and ratio scales, and apparent high dimensionality in metric multidimensional scaling solutions, can be "explained" by transforming data with an exponential transformation. A similarity approach to the former result was the subject of the previous section; we shall briefly give a partial explanation of the latter result in this section.

The usual interpretation of the apparent need for the exponential transformation is that subjects "underestimate large distances" (see Coombs, op. cit.); it occurs primarily with prothetic continua. But this is the Procrustean syndrome at its worst: what is really meant is that Euclidean distance (as the

distance function usually is in this context) overestimates subjective dissimilarities on "large distances" - that is, it is a bad model. We shall therefore compare our dissimilarity models on prothetic continua with the Euclidean distance, for "big distances".

This comparison is not a well-defined one because of the fact that (as was proved in Chapter 1) the two models do not give the same dissimilarity ordering: what is "big" for Euclidean distance is not necessarily "big" for the present models. The most usual case of "big" distances for the Euclidean model is, however, when at least one of the two points is well away from the origin. Thus if we take the cases that $x, y, u, v, \in \mathbb{R}^n$ are such that

$$h_x + h_y > h_u + h_v \quad (5.26)$$

$$\text{or } h_{x+y} > h_{u+v} \quad (5.27)$$

(which are not equivalent) we shall not be too far wrong. They will include a proportion of "small" distances; and conversely, some "large" distances will not be included. But in this case, to compare the performance of the two models, we look at

$$R_1 = \frac{D_{12}(x,y)}{D_{12}(u,v)} \cdot \frac{M_2(u,v)}{M_2(x,y)} ; \text{ and}$$

$$R_2 = \frac{D_{22}(x,y)}{D_{22}(u,v)} \bigg/ \frac{M_2(u,v)}{M_2(x,y)} .$$

We use D_{12} and D_{22} merely for simplicity;

$D_{1\beta}$ and $D_{2\beta}$ would give less straightforward but not radically different results.

Now

$$R_1 = \frac{\frac{M_2(x,y)}{h_x + h_y}}{\frac{M_2(u,v)}{h_u + h_v}} \cdot \frac{M_2(u,v)}{M_2(x,y)} = \frac{h_u + h_v}{h_x + h_y} \quad (5.28)$$

and

$$R_2 = \frac{\frac{M_2(x,y)}{h_x + y}}{\frac{M_2(u,v)}{h_u + v}} \cdot \frac{M_2(u,v)}{M_2(x,y)} = \frac{h_u + v}{h_x + y} \quad (5.29)$$

so by assumptions (5.26) and (5.27) respectively,

$$R_1 < 1 \quad \text{and} \quad R_2 < 1,$$

or

$$\frac{D_{12}(x,y)}{D_{12}(u,v)} < \frac{M_2(x,y)}{M_2(u,v)} \quad \text{and} \quad \frac{D_{22}(x,y)}{D_{22}(u,v)} < \frac{M_2(x,y)}{M_2(u,v)} \quad (5.30)$$

In other words, the present models, in these particular cases, predict the "underestimation" of "large" distances. This suggests they are better models for prothetic spaces, and that if used in multidimensional scaling, will give more valid, perhaps lower dimensioned, results.

Note again that this "underestimation" is not predicted for the equal-measure metathetic case.

5.2.2 Ekman's Law for metathetic continua

Ekman (1961) has suggested a subjective analogue of Weber's law, for prothetic continua:

$$\Delta x = \alpha x + \beta \quad (5.31)$$

where Δx is the j.n.d. on the continuum at the value x , and α and β are constants with β very near to zero. Stevens (1966) has called it "Ekman's Law" and suggested that decreasing sensitivity (increase in size of j.n.d.'s) up a scale was a characteristic of prothetic scales.

A similarity derivation of Ekman's Law was reproduced in section 2.3.1 of Chapter 2; it assumed that "a j.n.d. is constant on the similarity continuum". Once again, this can be applied to the metathetic case of the present models, giving

$$\alpha' = \beta(x, x+\Delta x) = F(\Delta x) \quad (5.32)$$

where α' is a constant and F is F_1 or F_2 of equations (5.4) and (5.5). In either case, since both are one-to-one, we have

$$x = F^{-1}(\alpha')$$

or

$$\Delta x = \alpha \quad (5.33)$$

where $\alpha = F^{-1}(\alpha')$, a constant.

In other words, the "equal similarity of j.n.d.'s" assumption also predicts constant sensitivity (constant discrimination, constant size of j.n.d.'s) on equal-measure metathetic continua, which is the usual approximate empirical result.

CHAPTER 6.APPLICATION TO SOME DATA AND TO MULTIDIMENSIONAL
SCALING

Up until this point, the development of the (dis)similarity models has been largely abstract and theoretical. In this chapter we shall look at a limited number of experiments which seem to be illustrative of the distinctions drawn within the models, and their practical implications. This will lead into a discussion of multidimensional scaling (MDS) with regard to the new models. Finally we shall briefly examine some empirical ordinal results, found by various workers in the field of similarity, as to whether they are consistent with the models of this thesis.

It will be noted that, except for the last section (ordinal results), all similarity data used in this chapter will be pooled data. Pooling over subjects is, quite rightly, a highly contentious procedure - see, for example, Carroll (1972), Shepard (1964), or the references of section 6.3 of this chapter. Individuals will have different sets of dimensions that they use, different weights on those dimensions, and possibly different judgment strategies and/or models. The hope in averaging data is that the pooled data will represent an

"average subject". The risk is that pooling will in fact create data that appears to fit a model that in fact fits none of the individual subjects. Shepard (op.cit) gave an example where pooled, but not individual, data indicated non-convex isosimilarity contours. Micko and Fischer (1970) and Fischer and Micko (1972) have shown that averaging Minkowski distances is invalid. We use pooled data here mainly because virtually all data published by Content model workers is pooled data; however we do think the "average subject" view does have some validity, especially at the "gross features" level we are largely aiming at, at present. We shall have something more specific to say on this in section 6.4.

6.1 SOME SIMILARITY DATA

6.1.1 In one dimension

The metathetic continua we are dealing with are our socalled "equal measure" continua (see Chapter 3). It was noted in section 3.2.2 of Chapter 3 that a very similar "equal measure" concept was used by many Content model workers to simplify their models, and that therefore a large body of similarity data for this latter type of space existed. It seems reasonable to look at some of this data to see whether it coincides with our metathetic continua.

We have therefore chosen some fairly typical examples of this type of experiment: from Ekman (1965) (which itself uses data reported in other papers). The data is presented in Table 6.1.

In each case, subjects were instructed to ignore any quantitative differences or "differences in intensity." Similarity estimates were given on a scale ranging from 0 ("no similarity") to 100 ("identity").

Ekman analysed the similarities by using the model

$$\cos\phi_{xy} = \frac{g(x,y)}{4} (g(x,y) + \sqrt{8+g^2(x,y)})$$

which is the solution for $\cos\phi_{xy}$ of the model of Ekman, et. al. (1964). Ekman factor analysed the cosines so obtained, by the method of principal factors, and rotated to simple structure.

In experiment 1, the stimuli were the nine Swedish words meaning in English: happy, pleased, content, sad, gloomy, depressed, agitated, impatient, restless. Ekman found three factors, which he interpreted as "sadness", "happiness", and "agitation". In fact, Ekman's scaling produced three clusters of three virtually indistinguishable points. There therefore seems no reason why more than two dimensions should be necessary, and this is confirmed in a similar

TABLE 6.1SIMILARITY DATA FROM EKMAN (1965)

Labelling refers respectively to the stimuli detailed
in the text.

Experiment 1.

	1	2	3	4	5	6	7	8	9
1									
2	.82								
3	.83	.90							
4	.05	.06	.09						
5	.05	.05	.07	.87					
6	.05	.06	.08	.86	.88				
7	.14	.11	.12	.47	.47	.47			
8	.19	.12	.12	.27	.33	.37	.65		
9	.20	.16	.09	.39	.38	.40	.80	.80	

Experiment 2.

	1	2	3	4	5	6
1						
2	.64					
3	.48	.84				
4	.42	.68	.76			
5	.37	.50	.39	.79		
6	.24	.42	.47	.64	.77	

Experiment 3.

	1	2	3	4	5	6	7
1							
2	.71						
3	.51	.82					
4	.40	.67	.91				
5	.29	.62	.89	.91			
6	.10	.29	.53	.70	.77		
7	.06	.10	.37	.60	.61	.94	

Experiment 4.

	1	2	3	4	5	6	7	8	9	10
1										
2	.80									
3	.63	.40								
4	.55	.60	.74							
5	.91	.73	.40	.66						
6	.19	.23	.27	.42	.06					
7	.02	.00	.11	.04	.00	.24				
8	.09	.08	.26	.17	.04	.81	.47			
9	.44	.51	.51	.67	.40	.66	.06	.44		
10	.88	.72	.73	.71	.84	.15	.05	.08	.39	

experiment by Lund (1975).

The stimuli of experiment 2 consisted of six mixtures of amyl acetate and n-heptanal, which were sniffed by the observers. They were in concentrations of 100, 87.5, 75.0, 50.0, 25.0 and 0.0 percent of amyl acetate. Subjective intensity was approximately constant for all stimuli. Two factors - one representing each chemical - were found.

Seven colour stimuli of constant brightness - nearly monochromatic light with transmission maxima at 522, 546, 560, 566, 570, 575 and 580m μ - were used in the third experiment. Again, two factors were found: "green" and "yellow".

The final experiment used ten words representing personality traits; in English translation: shy, inhibited, reserved, uncommunicative, timid, self-centred, independent, self-sufficing, unsociable, taciturn. The two factors found were interpreted as "Shyness" and "Self-centration".

If the present models are correct we would expect the similarity data to be well described by one of two transformations F_1 or F_2

$$F_1(x) = 1 - x, \quad 0 \leq x \leq 1 \quad (6.1)$$

$$F_2(x) = \frac{1-x}{1+x}, \quad 0 \leq x \leq 1 \quad (6.2)$$

of either a Minkowski distance (if the continuum is equal-measure metathetic) or one of the distance functions $D_{1\beta}$ or $D_{2\beta}$ (if the continuum is prothetic). Of course if the continuum is somewhere midway between prothetic and equal-measure metathetic, none of these distance functions will be exactly right.

A curve fitting (equally, a scaling) exercise was therefore carried out to see which, if any, of the six possible models fitted each set of data best in one dimension. This was done by minimising, over the n scale values $\{x_i\}_{i=1}^n$, the loss function

$$L = \sum_{i=1}^n \sum_{j=1}^n (F(\mathcal{S}(a_i, a_j)) - D(x_i, x_j))^2$$

where - there are n stimuli $\{a_i\}_{i=1}^n$;

the $\mathcal{S}(a_i, a_j)$ are the similarity data;

F is either F_1 or F_2 ;

D is D_1, D_2 or M .

Since $M(x_i, x_j) = |x_i - x_j|$, if $D = M$ then

$$\frac{\partial L}{\partial x_i} = -2 \sum_{j \neq i} F(\mathcal{S}(a_i, a_j)) \frac{|x_i - x_j|}{(x_i - x_j)} + 2 \sum_{j \neq i} (x_i - x_j)$$

so that for a minimum, if we take $\sum_{j=1}^n x_j = 0$
(as we may), we have

$$x_i = \sum_{j=1}^n F(\beta(a_i, a_j)) \frac{|x_i - x_j|}{x_i - x_j}.$$

Since the $\frac{|x_i - x_j|}{x_i - x_j}$ are determined simply by the order of the x_i , which can be found either by nonmetric unidimensional scaling or by unfolding methods, an analytic best-fit solution is easily found.

For D_1 or D_2 an analytic solution is not so easily found; here a steepest-descent iterative minimisation technique was used. In this case, best minima were found for $x_i \in \mathbb{R}^+$, $i = 1, \dots, n$, so it is immaterial whether D_1 or D_2 is the distance function used (we are therefore really testing four, not six models). We can, and shall, therefore use the notation $D = D_1 = D_2$ in this section. However, there were noticeable problems with local minima; it is hoped these were largely over come by use of several initial configurations. Even then, somewhat different configurations with nearly equal L-values were found, but these configurations did not differ enough to be of real practical significance.

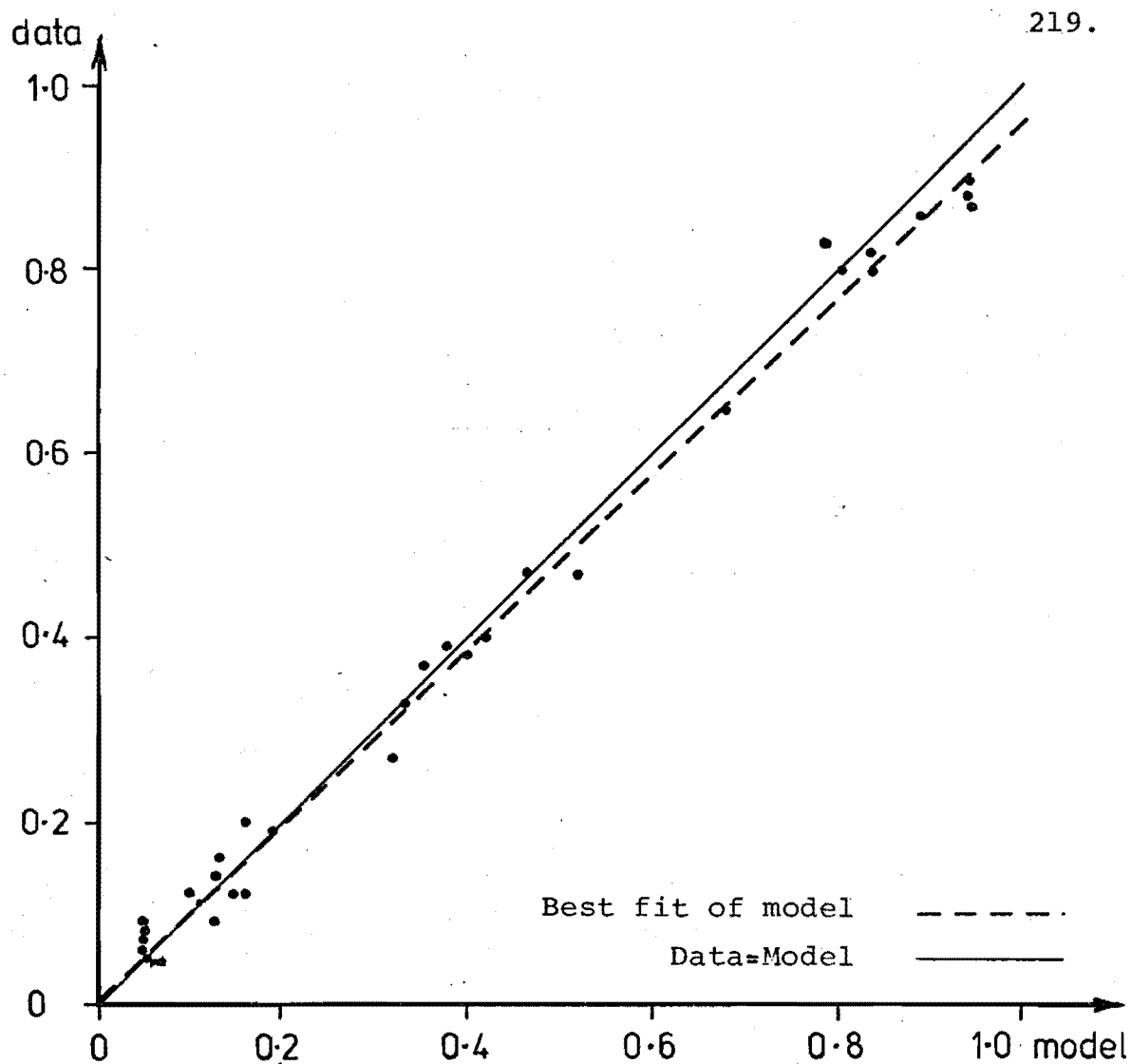
The minimum L-values occurred as follows.
 For experiments 1, 2 and 4, the best model was $F_2(D)$ - the ratio model of unidimensional similarity.
 For experiment 3 it was $F_1(M)$. However the fit for $F_2(M)$ in experiment 2 ($L = 0.0438$) was not much worse than that for $F_2(D)$ (0.0362). A linear regression of the theoretical similarity values versus the empirical values gave F-ratios that were highly significant (beyond the .005 level in all four cases). Experiment 2 has a poorer fit than the other three, though it is still highly significant. The results are tabulated in Table 6.2.

TABLE 6.2
Fit of best-fitting models for data in
Ekman (1965).

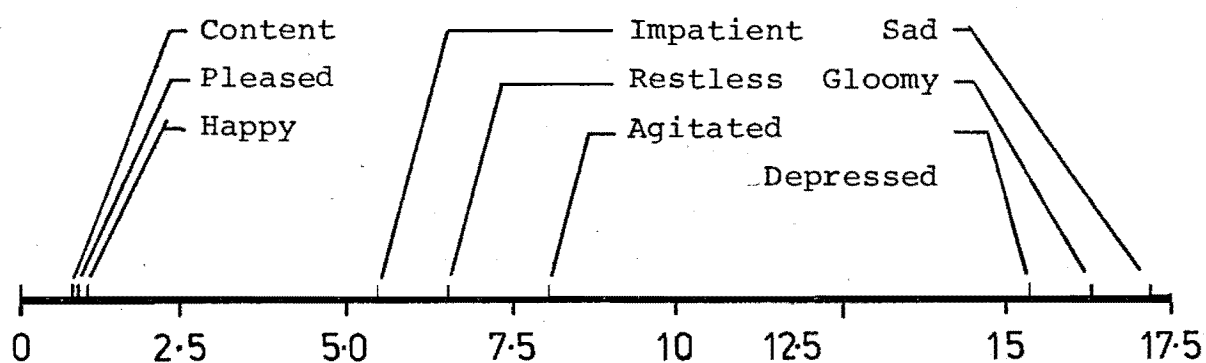
Expt.	Model	Gradient	L	F-Ratio	d.f.
1	D	F_2	0.0399	3623.8	25
2	D	F_2	0.0362	68.2	7
3	M	F_1	0.0503	443.4	12
4	D	F_2	0.320	300.3	33

d.f. = $(n(n-1)/2) - n - 2$ for n stimuli

The linear fit is plotted in Figures 6.1 - 6.4 respectively, along with the four scalings of the stimuli.

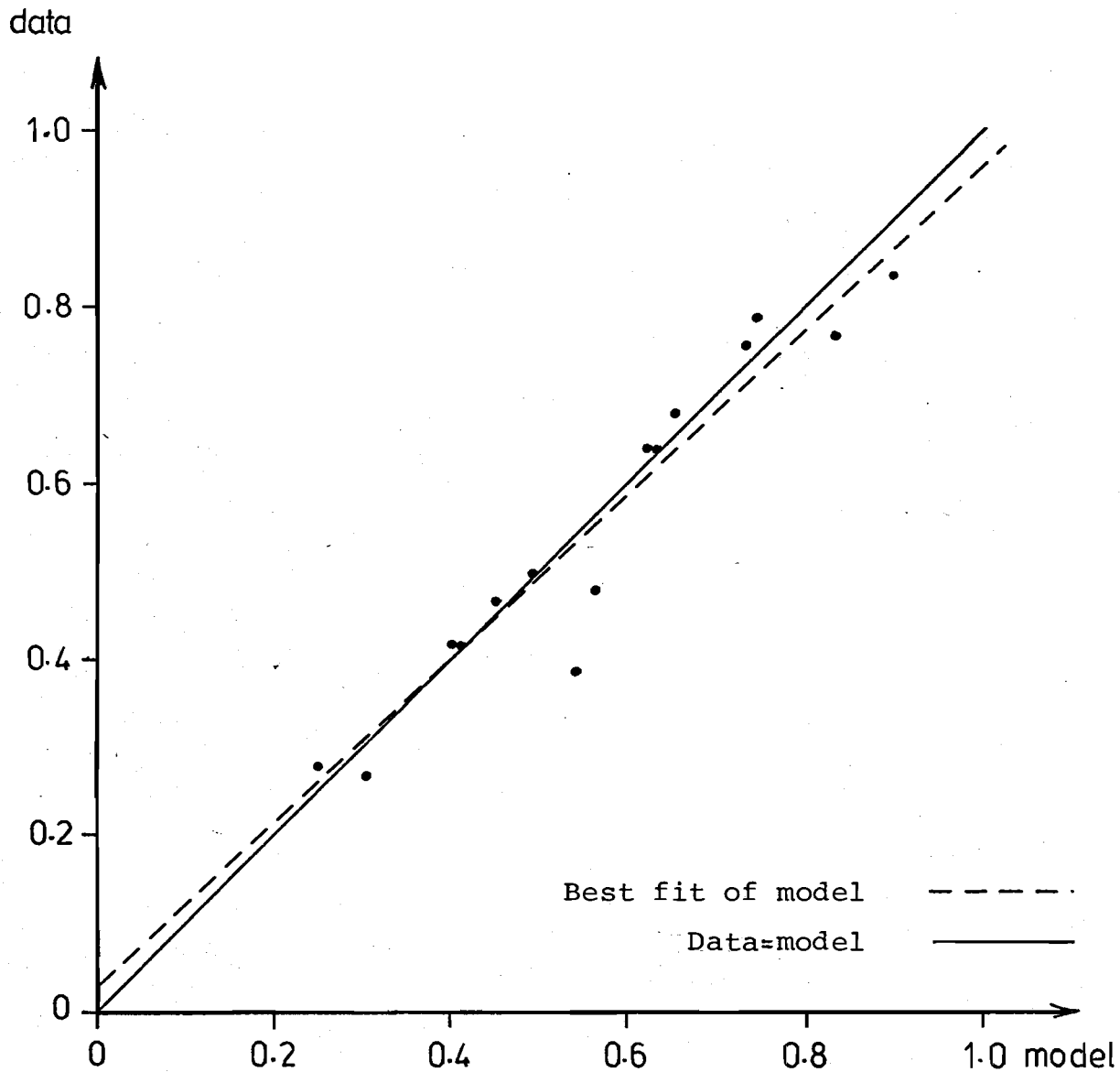


(a) Fit of model $F_2(D)$ to data.

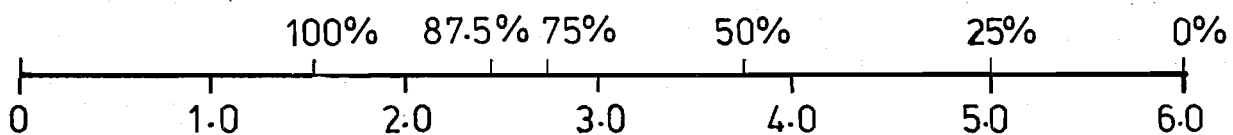


(b) Scaling of data.

FIGURE 6.1 Fit and scaling of data from Ekman (1965), Expt. 1.

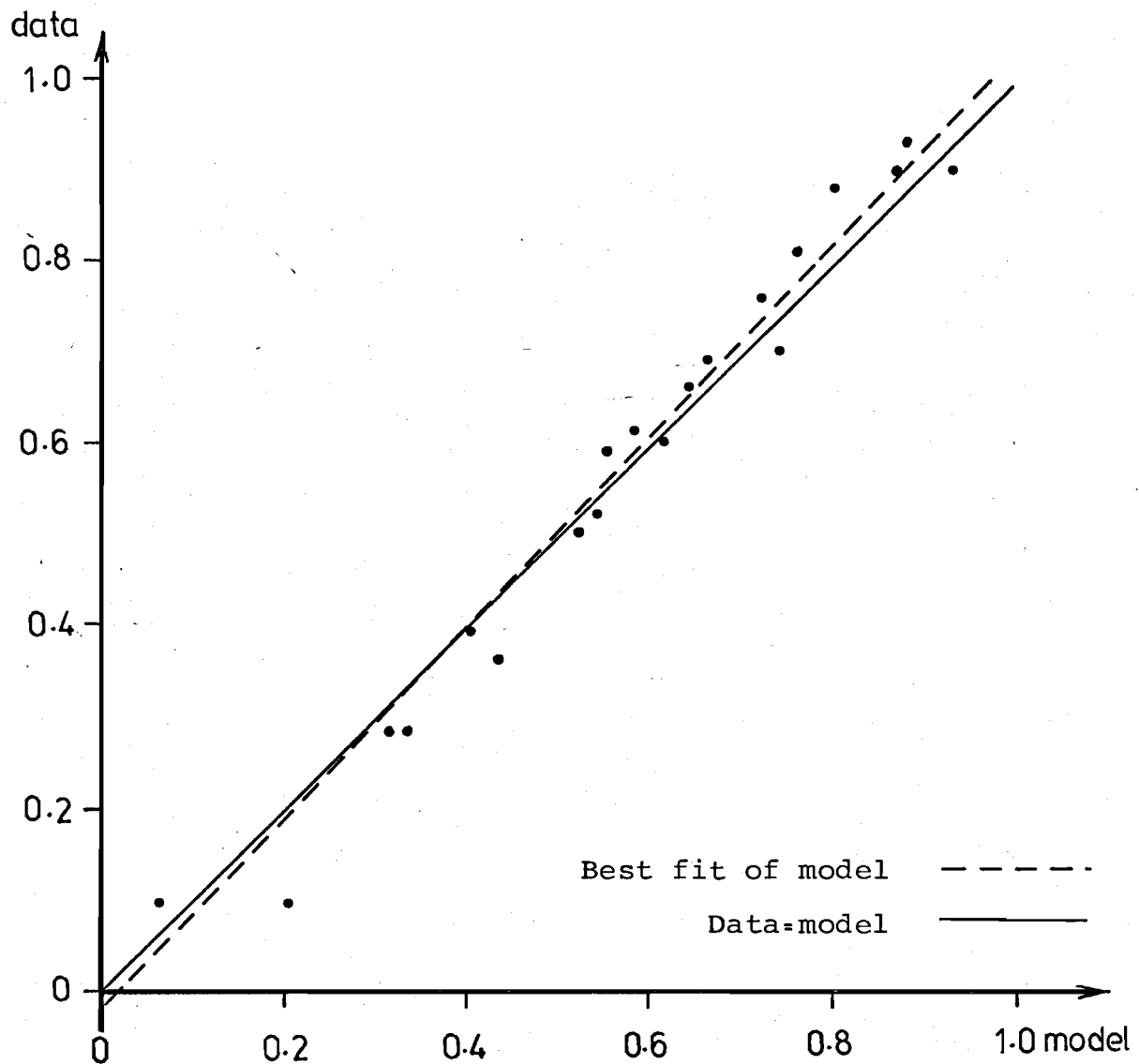


(a) Fit of model $F_2(D)$ to data.

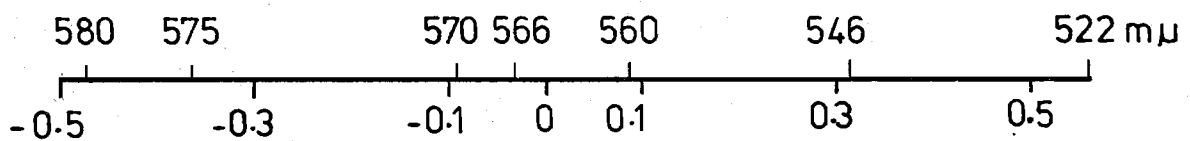


(b) Scaling of data: stimuli are marked in percentage amyl acetate.

FIGURE 6.2 Fit and scaling of data from Ekman (1965) Experiment 2.

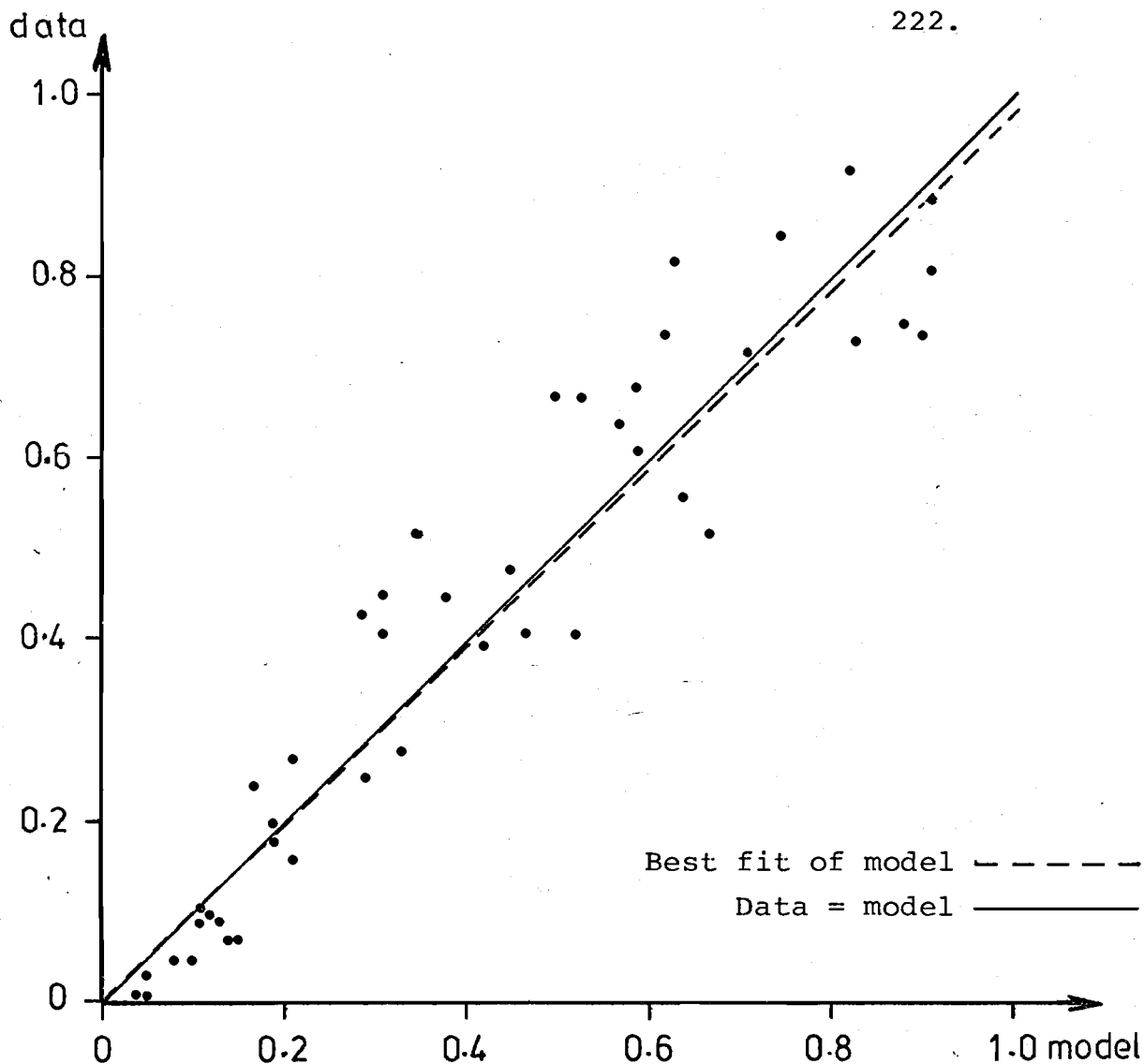


(a) Fit of model $F_1(M)$ to data.

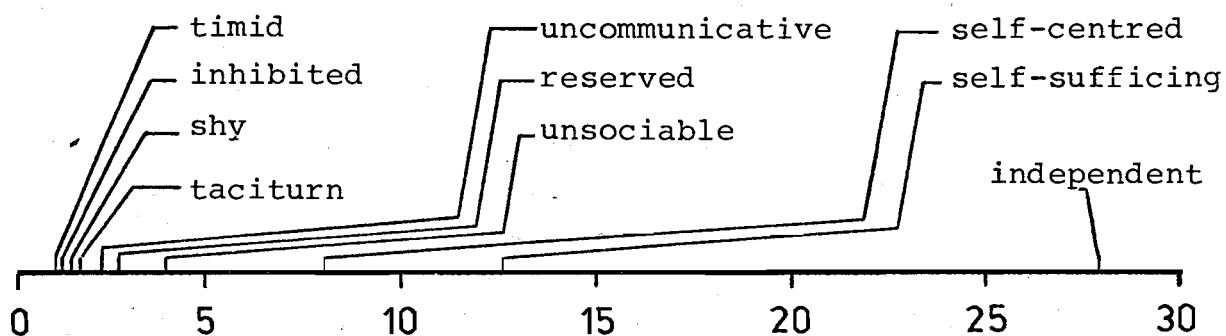


(b) Scaling of data: stimuli identified by their wavelengths in $m\mu$.

FIGURE 6.3 Fit and scaling of data from Ekman (1965) Experiment 3.



(a) Fit of model $F_2(D)$ to data.



(b) Scaling of data.

FIGURE 6.4 Fit and scaling of data from Ekman (1965)
Experiment 4.

We can interpret the results as follows.

Experiment 1 is on the prothetic continuum of sadness.

It should be noted that since $D\left(\frac{1}{x}, \frac{1}{y}\right) = D(x, y)$ for one dimension, an equivalent (in the sense of equally well-fitting) solution exists on a "happiness" continuum which would simply be a multiple of the reciprocal of that for sadness (this is consistent with an observation of Torgerson, 1960). A similar comment applies to both experiments 2 and 4. Recall that Ekman found three dimensions in this experiment and Lund, two. Thus the present model reduces the dimensionality while preserving an excellent fit. The subjects would appear to be using only one of the several latent dimensions of the stimulus set.

Experiment 2 can be described most simply as on the prothetic continuum of n-heptanal concentration or intensity (or, alternatively, amyl acetate concentration). The scaling in fact almost equally spaces the multiples of 25.0% concentration although the stimulus with 97.5% amyl acetate seems slightly displaced. But in view of the near equal fit of the $F_1(D)$ and $F_2(M)$ models, it would seem likely that this continuum is one which is midway between prothetic and equal-measure metathetic. This suggestion is in agreement with a very similar finding by Eisler (1963), and would to some extent explain results like those of Mitchell (1971), who found that odour intensity could

not be accurately described by a simple power law, giving reason to doubt the purely prothetic nature of this continuum. Mitchell (op.cit., p.3-25) also questions the use of grouped data from odour similarity experiments, providing data on the reliability of individual subjects.

Experiment 3 implies a metathetic continuum of hue in the green-yellow zone of the spectrum going from green to yellow (or vice versa). Superficially this might seem at variance with the more complex scale one would expect from one of the usual three-colour theories of colour vision (see for example, Judd, 1960, p.832ff.; Graham, 1965, Chapters 13 and 15; Gregory, 1966, p.121). However, the stimulus set is largely situated between the usual values given for the primaries green and red: around 550m μ and 600m μ respectively; and in this part of the spectrum, the primary of blue or violet has negligible effect and is nearly constant. Thus the main effect could be interpreted physiologically as a simple substitutive process between red and green sensitive cones (see also Indow, 1974b). The uneven spacing of the stimuli on the subjective scale (see Figure 6.3(b)) also strongly suggests that the scale is, at least approximately, a constant discrimination one. In

particular, it predicts a maximum in discriminability (a minimum j.n.d.) for wavelengths between 570 and 575m μ . This is in accord with data found by direct methods (see Graham, 1965, Chapters 12 and 15; and Gregory, op.cit., p.124). This is also in accord with the suggestion under "Ekman's Law" (section 5.2.2) in Chapter 5 that equal-measure metathetic scales could be constant discrimination ones.

Finally, experiment 4 produces a prothetic scale which seems best interpreted as a "self-confidence" scale.

These four analyses show that the conventional content model interpretation of "equal measure" (i.e. equal vector lengths of the stimulus points) does not coincide with that of the present models. However the present models appear to fit well, and their approach to the prothetic versus metathetic distinction has some independent validation.

Two points should be noted. Firstly, the present models seem to quite effectively reduce the dimensionality required to explain the data. This implies that subjects do not use all of the, sometimes quite distinctive and therefore perceived, dimensions latent in the interstimulus variations exemplified in the stimulus set, when judging similarities.

Fenker and Brown (1969) and Green and Carmone (1971) are among the many who previously have recognised this type of behaviour.

Secondly, the simple ratio model of unidimensional similarity seems to have been revived by these analyses; this may be because we are now definitely working in subjective space.

6.1.2 In two dimensions

From both a similarity modelling and a MDS point of view, one of the most interesting experiments carried out using similarity judgments is that in Eisler (1967). This appears to be a critical one in determining the meaningfulness of various methods of analysing such experiments. As was mentioned in Chapter 2 (section 2.2), Eisler's experiment, along with his own analysis and that done by Roskam (1972), provides an example of how two different analyses (i.e. two different models) can give two equally interpretable but different solutions. One of these solutions must of course be wrong, so interpretability is no guarantee of validity.

Eisler's experiment is detailed in Eisler (1967); each stimulus consisted of two points (produced by light shining through holes in a metal sheet), the bottom left hand one of which was always fixed, and the other taking a position in a 4x4 grid (see Figure 6.5).

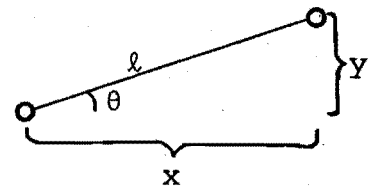
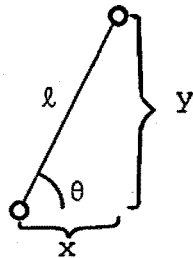
CO GO KO OO

BO FO JO NO

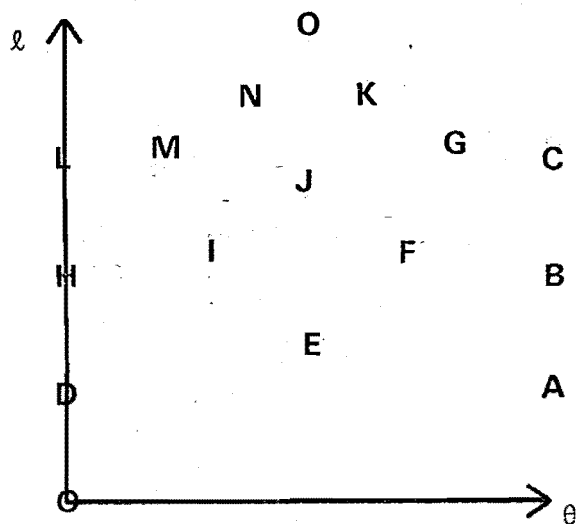
AO EO IO MO

O DO HO LO

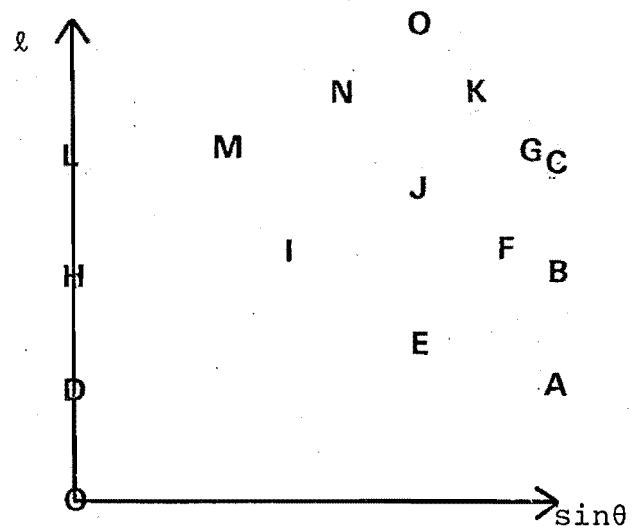
(a) The original design of light holes: the unlabelled hole was lit for all stimuli. The labelled holes form the Eisler configuration in terms of the horizontal and vertical axes.



(b) A typical stimulus pair (the labelling and line were of course not present).



(c) The Eisler configuration in terms of angles and lengths.



(d) The Eisler configuration in terms of sines of angles and lengths.

FIGURE 6.5 Eisler's Experiment.

Subjects made similarity judgments between all pairs of such stimuli, on a ten-point scale. The arithmetic mean was taken of these ratings, and they were divided by 10 to give similarity values between 0 and 1 (see Table 6.3).

The reason why this experiment is an interesting one is that there are at least two obvious possible descriptions of each stimulus. It can be described in terms of the projections the variable point of the stimulus makes on the horizontal and vertical axes; or it can be described in terms of the length of the line joining its two points, and the angle (or some function of the angle) that line makes with the horizontal axis. Other such descriptions naturally exist, as will be seen. The descriptions are themselves completely equivalent mathematically in that any one can be calculated from any other; but the interesting point in the experiment is which (if either) of the two above descriptions is used by the observer to make judgments, since each description will give rise to a different set of similarities.

A good MDS solution should provide up to two dimensions that are interpretable in terms of one of the possible descriptions. It would then be assumed that this pair of dimensions is the one used by the subject. It was pointed out in section 2.3 of

TABLE 6.3

SIMILARITY DATA FROM EISLER (1967)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
A															
B	.677														
C	.506	.773													
D	.443	.255	.171												
E	.633	.474	.304	.455											
F	.479	.803	.662	.232	.666										
G	.315	.646	.863	.136	.402	.748									
H	.328	.407	.297	.708	.454	.418	.304								
I	.394	.509	.368	.333	.791	.670	.397	.602							
J	.308	.590	.583	.209	.638	.861	.731	.387	.687						
K	.252	.519	.723	.136	.458	.750	.902	.253	.483	.745					
L	.258	.330	.410	.564	.280	.379	.409	.811	.501	.480	.381				
M	.252	.387	.394	.270	.535	.576	.477	.493	.830	.703	.510	.627			
N	.246	.436	.544	.179	.542	.681	.659	.345	.682	.875	.785	.438	.760		
O	.181	.411	.590	.117	.474	.634	.792	.240	.483	.775	.916	.342	.490	.789	

Chapter 2 that interpretability of dimensions in this sense is often taken as an indication of the suitability to the data of the model used in the MDS algorithm.

In the present case however, two different models used for MDS gave the two different sets of dimensions. In his original report, Eisler used a vector-type Content model to scale his data; he found that the two dimensions of his scaling quite clearly corresponded to the vertical and horizontal axes of the stimulus space (he also found two other dimensions which were rather difficult to interpret). Roskam (1972) and Eisler both analysed the data using several Minkowski-based non-metric MDS algorithms. The result in every case was highly interpretable: the axes corresponded to stimulus dimensions of line length and some function of angle (Roskam suggested the function was the sine function; Eisler, less convincingly, suggested it was the identity function).

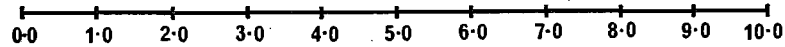
The result therefore casts doubt on the validity of using the results of any type of MDS to deduce the dimensions used by observers in making similarity judgments. Two approaches suggest themselves to check the nature of this particular result. Firstly, this experiment was designed to give a grid pattern when the physical dimensions are the horizontal and vertical axes - we shall henceforth call this design

"Eisler's configuration"; this type of design allows very easy checking of the interpretation of the axes. A second experiment using the same type of stimuli but with a design that gives a grid pattern in the physical lengths-versus-sines space would check whether the result is merely an artifact of the particular experimental design.

Secondly, one could use the method of Monte Carlo studies by taking the type of configuration used in the experiment, calculating the interpoint distances under a variety of models, and scaling these distances with the different models used for scaling the empirical data. It would then be seen whether artifactual dimensions can result from using the wrong models for scaling. We shall discuss the use of our models in non-metric MDS and present the results of some Monte Carlo studies in section 6.2. In this section we present the results of an auxiliary experiment as suggested above.

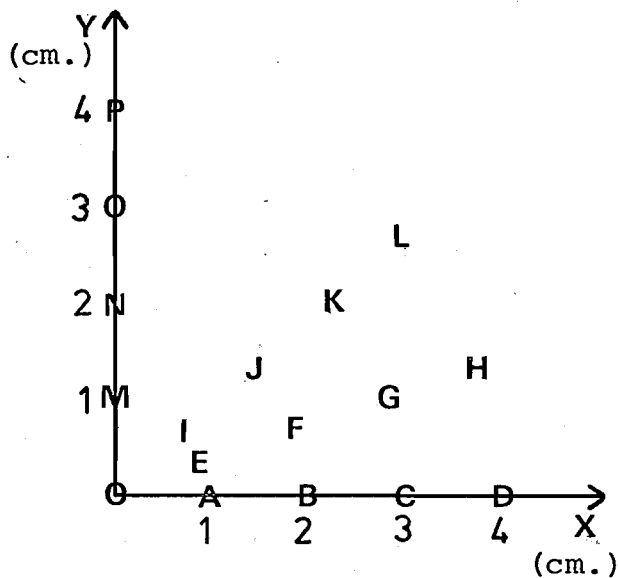
6.1.2.1 Stimuli

Instead of using the light-holes arrangement devised by Eisler, we simply used black dots on white paper, the bottom lefthand one being fixed with respect to the similarity scale used by subjects to mark judged similarities (see Figure 6.6(a)).

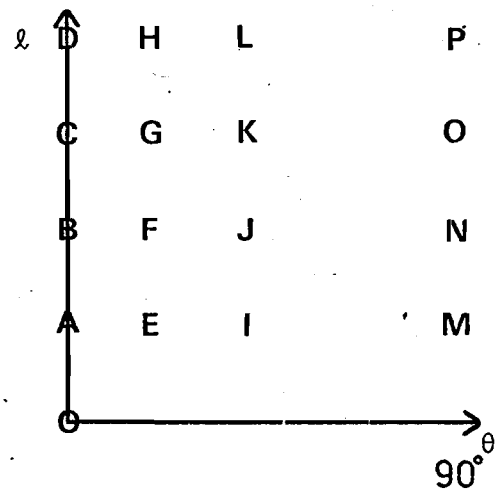


Similarity Scale

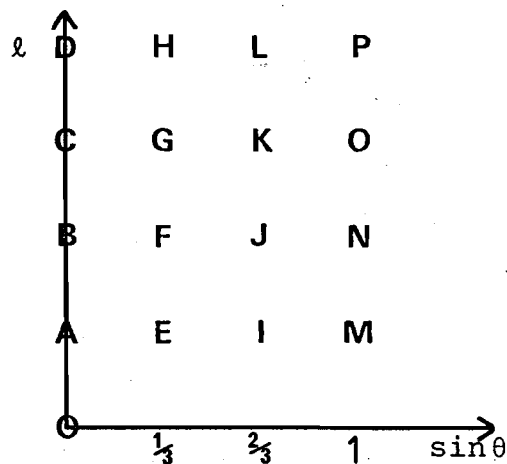
(a) A typical stimulus pair presentation (actual size).



(b) Design in terms of horizontal and vertical components.



(c) Design in terms of angles and lengths.



(d) Design in terms of sines and lengths.

FIGURE 6.6 Experimental Design (see Fig. 6.5 for notation).

The stimuli (16 in all) were chosen to give a 4x4 grid in the physical space of lengths versus sines (see Figure 6.6(b), (c)). All pairs of such stimuli, including all sixteen pairs of identical stimuli, (136 pairs in all) were made up into a booklet in a random order, three pairs to a page. Each pair of stimuli was therefore presented only once to each subject, and the pairs were presented in the same order to all subjects.

6.1.2.2 Subjects

Subjects were 53 second year psychology students, naive to both practice and theory of psychometric experiments of this type. They took part in the experiment as part of their classwork for the year, although participation was not compulsory.

6.1.2.3 Instructions

Instructions to subjects were printed on the first page of the booklet of stimulus pairs; they read as follows:

This is an experiment about similarity. In this booklet there are figures made up of a pair of point patterns, one to the left and one to the right. Each pattern consists of two dots. Your task is to estimate the similarity between the two patterns on a scale, which you will find beneath each figure, between zero and ten. Zero stands for no similarity at all, and ten stands for identity. Make your estimates as precise as you can by marking the scale at the corresponding value.

Here are a few examples for practice.

6.1.2.4 Procedure

Subjects were asked to read the instructions and do three practice trials. They were then asked if they had any questions, and proceeded with the experiment. Two experiments were carried out simultaneously: the present one, and another similarity experiment, of identical form but with different stimuli, which is not presented here. (The stimuli in that experiment were pairs of brick walls, like those used by Gregson, 1974, whose two heights were proportional to the sine and length values respectively, of the dot stimulus experiment: for details, see the unpublished paper by D. Cargo, 1973.) Half the subjects did one experiment first, half the other.

6.1.2.5 Results

To be consistent with Eisler's treatment of his data, similarity values were averaged over subjects. Of the 53 subjects, four were not included in the average because it was obvious from inspection of their judgments that they were not adhering to the instructions in a consistent way. One used dissimilarities, two showed little recognition of the meaning of the identity judgment, and the remaining one used only the integers of the lowest part of the scale, making frequent use of zero. The lower triangle of this matrix of average similarities, including the main diagonal of "identity" judgments, and the lower triangle of the

TABLE 6.4

DATA FROM DOT STIMULUS EXPERIMENT

(a) Matrix of average similarities.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
A	.977															
B	.749	.975														
C	.629	.780	.978													
D	.578	.740	.914	.991												
E	.723	.539	.485	.368	.980											
F	.571	.702	.614	.474	.725	.975										
G	.414	.537	.622	.656	.671	.757	.959									
H	.373	.448	.516	.668	.578	.711	.823	.948								
I	.630	.470	.381	.363	.784	.625	.537	.440	.986							
J	.340	.583	.477	.444	.611	.790	.735	.624	.672	.981						
K	.327	.475	.592	.570	.427	.503	.683	.655	.570	.770	.963					
L	.272	.411	.426	.573	.385	.491	.551	.618	.481	.627	.796	.963				
M	.525	.465	.321	.239	.560	.497	.510	.326	.667	.463	.314	.264	.986			
N		.454	.454	.304	.356	.477	.461	.361	.376	.599	.525	.378	.707	.984		
O	.215	.346	.491	.449	.244	.346	.463	.488	.267	.503	.513	.516	.580	.733	.971	
P	.189	.244	.293	.468	.231	.290	.359	.408	.245	.366	.391	.583	.467	.615	.776	.984

TABLE 6.4, continued

(b) Matrix of corresponding variances.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
A	.003															
B	.021	.005														
C	.041	.035	.004													
D	.052	.025	.009	.001												
E	.045	.044	.048	.045	.003											
F	.047	.052	.042	.057	.021	.005										
G	.048	.055	.061	.047	.034	.026	.008									
H	.053	.043	.047	.049	.046	.025	.038	.008								
I	.047	.054	.039	.042	.039	.036	.045	.045	.001							
J	.037	.057	.036	.045	.040	.031	.036	.031	.042	.004						
K	.034	.050	.052	.042	.043	.044	.042	.035	.037	.024	.006					
L	.036	.048	.052	.049	.044	.040	.038	.054	.058	.044	.024	.010				
M	.079	.062	.056	.048	.065	.056	.044	.039	.048	.054	.038	.035	.001			
N		.084	.076	.048	.044	.060	.050	.038	.037	.050	.046	.044	.038	.001		
O	.044	.051	.077	.091	.036	.046	.059	.064	.032	.040	.066	.055	.038	.025	.004	
P	.042	.040	.051	.079	.037	.033	.046	.073	.042	.040	.062	.051	.045	.031	.028	.001

corresponding variances are given in Tables 6.4(a) and (b) respectively.

An error in the experimental design should be noted. The pair of stimuli which should have been N and A were in fact B and A, so data for the pair (N,A) is missing. However, this provides a small consistency check on the experiment. The two mean judgments which should be the same were 0.772 (standard error 0.018) and 0.726 (s.e. 0.023). In the analysis following, $\bar{g}(A,B)$ was taken to be the mean of these two values, 0.749, and $\bar{g}(N,A)$ to be missing data.

An indication of the reliability of the results can be found from a comparison of the first half of the 49 subjects (who did the reported experiment first) and the second half (who did D.Cargo's experiment first). A linear regression was therefore performed between the two halves; this gave an F-ratio of 453 with an intercept on the second-half (y) axis of -0.065 and a slope of 0.962. The difference between the two appears to be insignificant.

6.1.2.6 Analysis

The similarity data was scaled using Young's (1973) POLYCON non-metric MDS algorithm using Euclidean and Cityblock models, in from one to three dimensions.

The most interesting solutions, on both a priori and goodness of fit grounds, are those in two dimensions. The best fit for these solutions was for the Cityblock scaling with a stress (Kruskal's formula) of 11%; this compares well with the comparable result of Roskam's, scaling Eisler's data. It gave a solution that was a quite recognisable grid, though somewhat distorted in (what were physically) its outermost points. The Euclidean two dimensional configuration (stress 16%) was more distorted but still recognisably a grid. The two two-dimensional solutions are presented in Figures 6.7 and 6.8.

A scaling like that used by Eisler in his analysis was not performed because it requires estimations of "vector lengths" (viz., h_x for a stimulus x); this seems to involve a very strong prior assumption of the dimensions to be used, since the "vector length" would vary with the specification of a stimulus used. It is also very difficult to conceive of a suitable way to instruct subjects to find the "vector length" when the "vector" has components of lengths and sines - it is obvious if the dimensions are the horizontal and vertical axes.

Considering only the Minkowski-based scaling, therefore, the result Roskam and Eisler found does not appear to be merely an artifact of the stimulus set

CITYBLOCK ANALYSIS OF DATA FROM ROSENBERG'S DOT STIMULUS EXPERIMENT.
DERIVED CONFIGURATION: DIMENSION 1 (X-AXIS) VERSUS DIMENSION 2 (Y-AXIS)

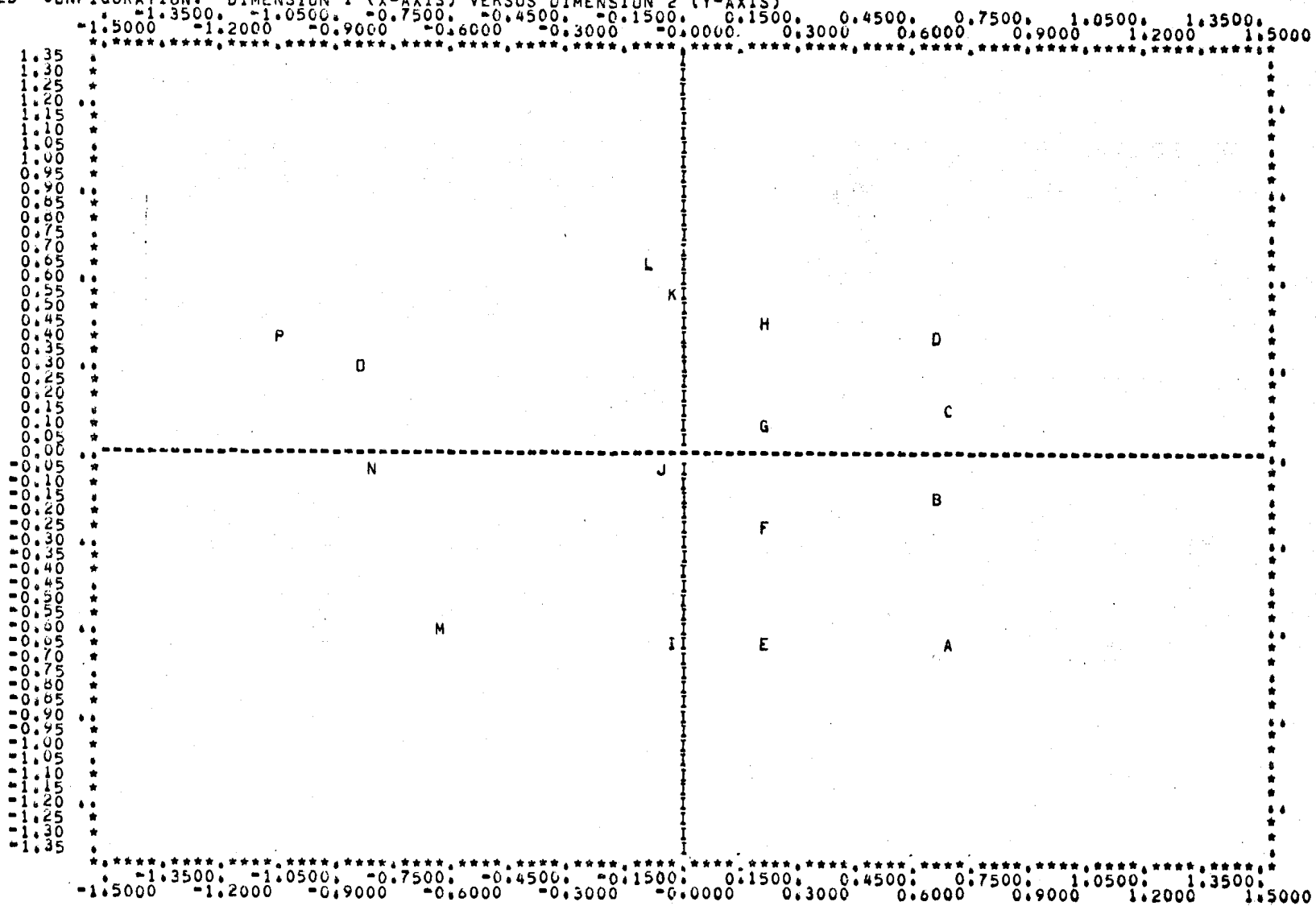


Figure 6.7: City-block analysis of dot experiment: stress = 11%

EUCLIDEAN ANALYSIS OF DATA FROM ROSENBERG'S DOT STIMULUS EXPERIMENT.
 DERIVED CONFIGURATION, DIMENSION 1 (X-AXIS) VERSUS DIMENSION 2 (Y-AXIS)

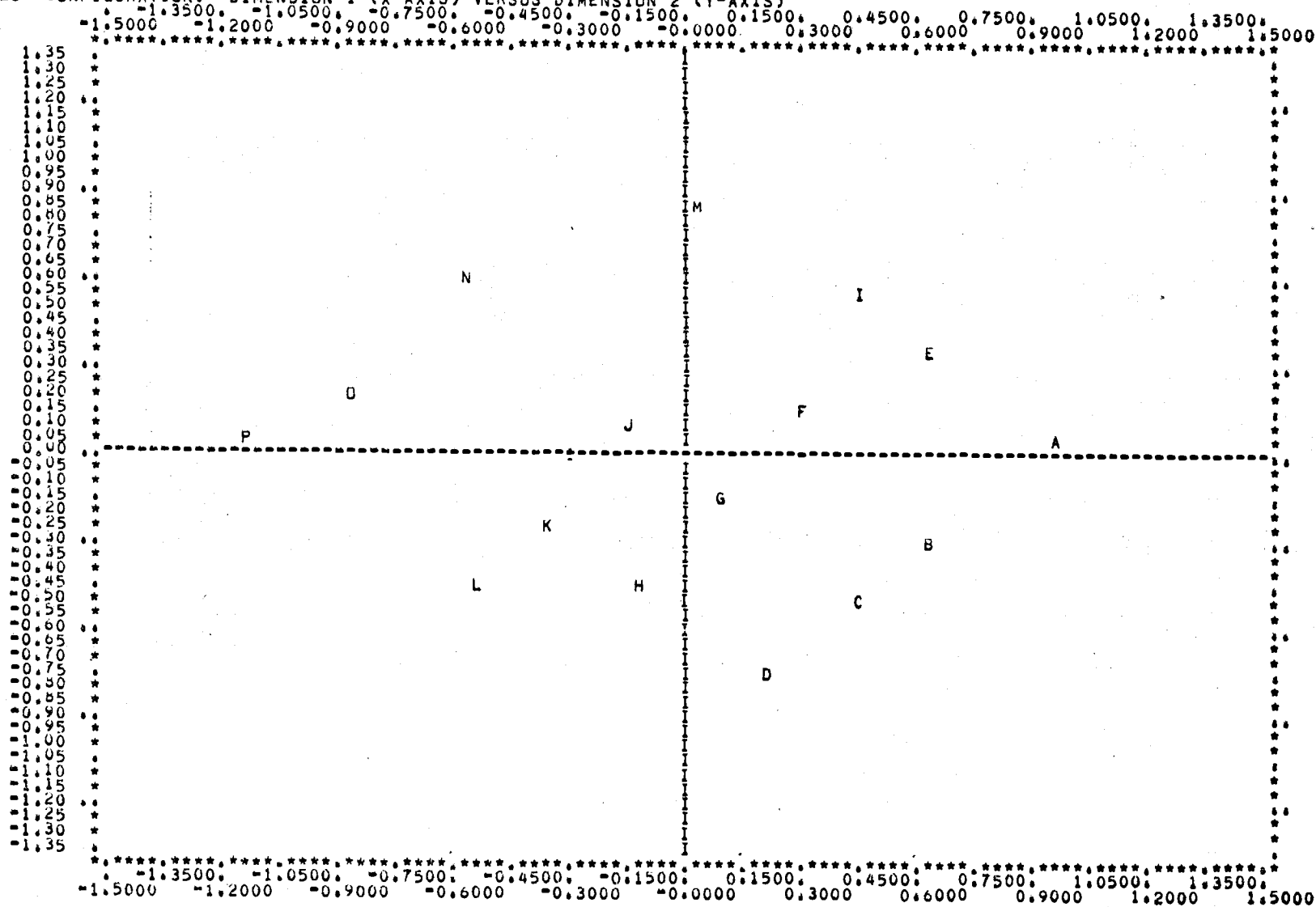


Figure 6.8: Euclidean analysis of dot experiment: Stress = 16%

used (though this possibility still cannot be completely ignored). A consistent interpretation of this type of stimulus, in terms of sines and lengths, seems possible. We must therefore look at the models inherent in the MDS techniques used to find an explanation of the apparent ambiguity in the stimuli.

6.2 MULTIDIMENSIONAL SCALING

We wish to find out whether scaling of data from one model, by a MDS algorithm based on a different model, can give artifactual but interpretable dimensions. The two classes of model we are concerned with here are the Minkowski distances and the normalised distances $D_{1\beta}$ and $D_{2\beta}$ discussed in previous chapters. Thus the first step would seem to be to construct a (preferably non-metric) MDS algorithm based on models $D_{1\beta}$ and $D_{2\beta}$.

6.2.1 Scaling with $D_{1\beta}$ or $D_{2\beta}$

The simplest way to construct a non-metric MDS algorithm with $D_{1\beta}$ and $D_{2\beta}$ would be merely to adapt an existing algorithm by changing the dissimilarity model on which it is based. Young's POLYCON (see Young, 1973, for a technical description of its operation) seemed a very suitable choice since it already has a wide range of options and is designed to make the introduction of new models a relatively easy task.

The two main requirements for modifying POLYCON are the derivatives, with respect to each of the point coordinates, of the distance functions, and a method for finding an initial configuration for the non-metric part of the program. The first requirement is easily met: for $x, y \in \mathbb{R}^n, \beta \in \mathbb{R}, \beta \neq 0, \beta' = (1-\beta)/\beta$,

$$\frac{\partial}{\partial x_i} D_{1\beta}(x, y) =$$

$$D_{1\beta}(x, y) \left[\frac{|x_i - y_i|^\beta}{\sum_k |x_k - y_k|^\beta (x_i - y_i)} - \frac{|x_i|^\beta}{x_i} \frac{(\sum_k |x_k|^\beta)^{\beta'} + (\sum_k |y_k|^\beta)^{\beta'}}{(\sum_k |x_k|^\beta)^{1/\beta} + (\sum_k |y_k|^\beta)^{1/\beta}} \right]$$

and

$$\frac{\partial}{\partial x_i} D_{2\beta}(x, y) =$$

$$D_{2\beta}(x, y) \left[\frac{|x_i - y_i|^\beta}{\sum_k |x_k - y_k|^\beta (x_i - y_i)} - \frac{|x_i + y_i|^\beta}{\sum_k |x_k + y_k|^\beta (x_i + y_i)} \right]$$

The second requirement - a procedure for finding an initial configuration - is far more difficult. In fact we could find no satisfactory answer. The main reason for this is that the stress function surface, under these distance functions, is pitted with local minima - far more even than is the case with Minkowski distances (see Arabie, 1973, for example). This was further confirmed, as reported above, even in the simpler version of the one-dimensional case. Thus the initial configuration

must be very good indeed (i.e. very close to the true minimum) for a satisfactory non-metric solution.

To find out just how close the initial configuration had to be to the best one, a simulation was run on POLYCON with the modification to allow $D_{1\beta}$ dissimilarities. This simulation consisted of scaling $D_{1\beta}$ dissimilarities from a given configuration, using as initial configuration the given configuration itself, but with random noise written onto it. Only at a very minor level of noise did the algorithm find the correct minimum stress solution. It therefore seems that if a solution is to be found, it will probably have to be a wholly analytic one (comparable to a Factor Analysis, for example), since any iterative procedure will make very little difference. Admittedly, only one type of iterative algorithm was tried, but since the problem apparently is the roughness of the stress surface, it seems unlikely that other types of algorithm would give much improvement.

No analytic scaling method for $D_{1\beta}$ and $D_{2\beta}$ is at all obvious, if only (dis)similarity data is allowed as input (if estimates of "vector lengths", $(\sum_{i=1}^n |x_i|^\beta)^{1/\beta}$, or "vector sums", $(\sum_{i=1}^n |x_i + y_i|^\beta)^{1/\beta}$, were allowed, then an algorithm like that used by

Eisler, 1967, would be quite straightforward to construct). We therefore investigated Eisler's algorithm instead. This seems justified for two reasons. Firstly, his similarity model

$$\beta(x,y) = \frac{2\min(h_x, h_y)}{h_x + h_y} \cos\phi_{xy}, \quad x,y \in \mathbb{R}^{n+} \quad (6.3)$$

is not unlike our models (see discussion in previous chapters). This was borne out in our Monte Carlo study. And secondly it was used by Eisler to analyse his data so artifactual results will be of substantial interest.

Eisler's algorithm is, briefly, as follows. His model (equation 6.3) implies that

$$h_x y_x \cos\phi_{xy} = \frac{\beta(x,y) (h_x + h_y)}{2\max(h_x, y_x)} \quad (6.4)$$

so that from similarity judgments $\beta(x,y)$ and estimates of vector lengths h_x, h_y , the lefthand side of (6.4), a scalar product, can be estimated. These scalar products are factor analysed using principal components, with the squared vector lengths as communalities.

6.2.2 Monte Carlo studies

Monte Carlo studies were carried out on both Eisler's algorithm and on POLYCON using Minkowski distances. In the former case both Euclidean and D_{12}

dissimilarities were calculated from the points of the Eisler configuration expressed either in terms of horizontal and vertical components, or in terms of sines and vector lengths. Since Eisler's algorithm requires similarities, the dissimilarities were transformed by each of the two similarity gradients proposed. The best factors from the varimax-rotated principal components analysis were rotated to either the vertical-horizontal or sine-length configuration (which of course is in two dimensions). It made little difference, to which of these configurations the factors were rotated. Only the M_2 and D_{12} models were used because M_1 , D_{11} , D_{21} and D_{22} , from experience in the POLYCON study, make only minor differences in this type of context.

With POLYCON, D_{11} , D_{12} , D_{21} , and D_{22} dissimilarities were calculated from both forms of the Eisler configuration, and were scaled directly. There was obviously little point, either in including Minkowski models in the study, or in transforming the dissimilarities to similarities, since POLYCON is a non-metric algorithm.

The design for each of the two algorithms is summarised in Table 6.5.

TABLE 6.5

Design for Monte Carlo Studies

Note: The "configuration used" is the Eisler configuration expressed in terms of projections on the horizontal and vertical axes (P) or sines and lengths (SL). For gradients, see equations (6.1) and (5.2).

(a) For Eisler's (1967) algorithm

Study	Dissimilarities model	Gradient	Configuration Used
1	D_{12}	F_1	SL
2	D_{12}	F_2	SL
3	D_{12}	F_1	P
4	D_{12}	F_2	P
5	M_2	F_1	SL
6	M_2	F_2	SL
7	M_2	F_1	P
8	M_2	F_2	P

(b) For POLYCON

All possible combinations of the following parameter values:

Parameter	Scaling Model	Dissimilarities Model	Configuration used	No. of Dimensions of Scaling
Values	M_1, M_2	$D_{11}, D_{12},$ D_{21}, D_{22}	P, SL	3, 2

Looking first at the results of the Eisler algorithm studies, if Eisler's similarity model, equation (6.3), is in fact very like our D_{12} model with either of the two gradients, then one would expect his algorithm to give a scaling which is nearly identical to the configuration used to generate the dissimilarity data under D_{12} (i.e. for studies 1-4). As figures 6.9 and 6.10 exemplify, this is so, although there is distortion in some cases, particularly with gradient F_2 . The results are close enough however, to be able to recognise the true dimensions.

The most interesting results are from studies 5 to 8, which use Euclidean (M_2) distances to generate dissimilarities. With dissimilarities calculated from the Eisler configuration expressed in sines and lengths, the scaling has a resemblance to this configuration, but is rather distorted (see, as example, Figure 6.11). With dissimilarities calculated from the Eisler configuration expressed as projections on the horizontal and vertical axes however, the scaling has very little resemblance to the original configuration, but it is interpretable in a quite different way. If Figure 6.12 (the scaling of study 8) is compared with Figure 6.13(a), the resemblance

CONFIGURATION FROM MONTECARLO STUDY 2, FROM EISLER CONFIGURATION, TO BE ROTATED TO THE EISLER CONFIG. IN TERMS OF POLARS.
 ROTATED CONFIGURATION. DIMENSION 1 (X-AXIS) VERSUS DIMENSION 2 (Y-AXIS)

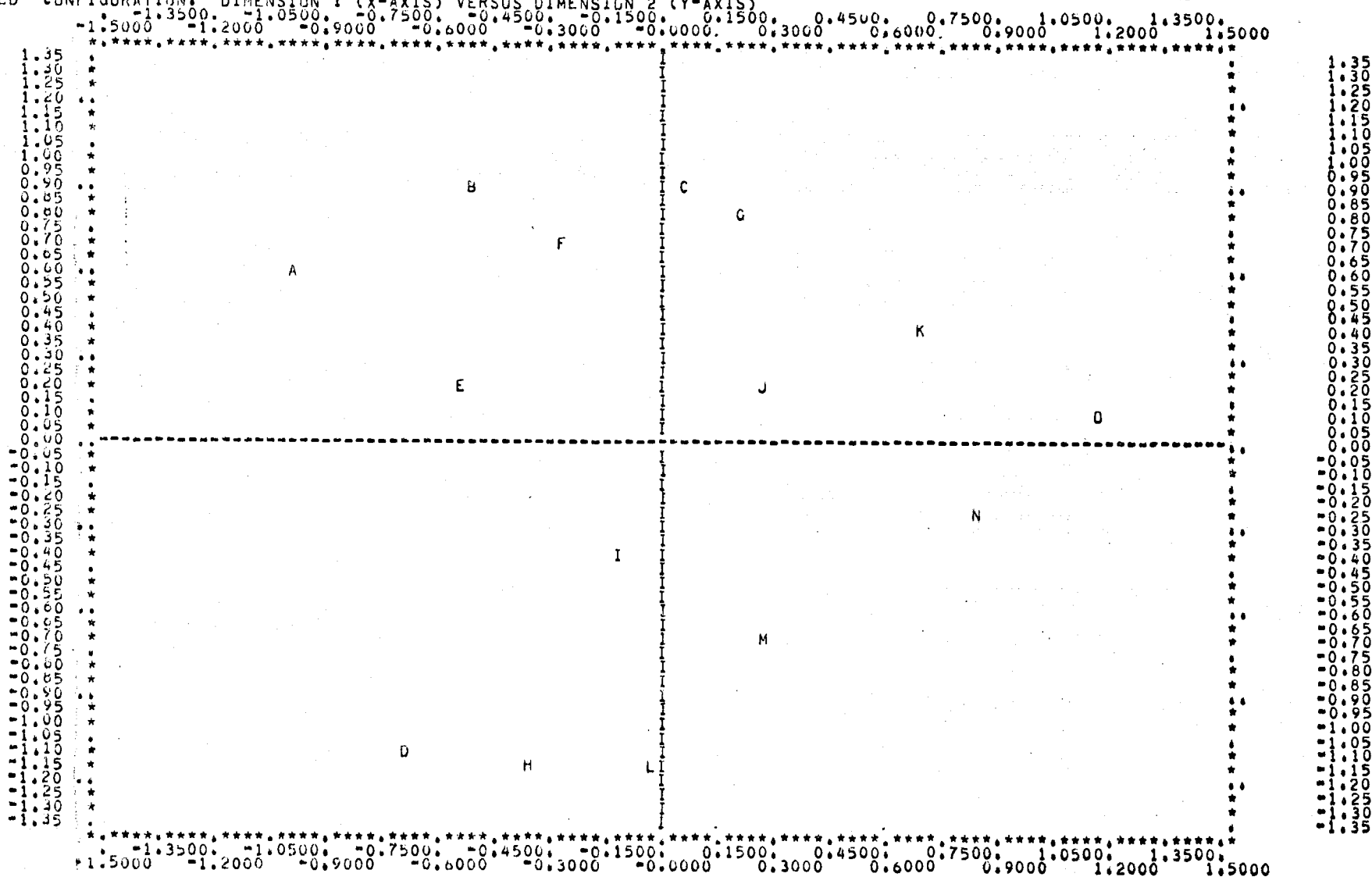


Figure 6.9: Monte Carlo Study 2 (see Table 6.5)

CONFIGURATION FROM MONTECARLO STUDY 3, FROM EISLER CONFIGURATION, TO BE ROTATED TO THE EISLER CONFIG. IN TERMS OF POLARS.
 ROTATED CONFIGURATION. DIMENSION 1 (X-AXIS) VERSUS DIMENSION 2 (Y-AXIS)

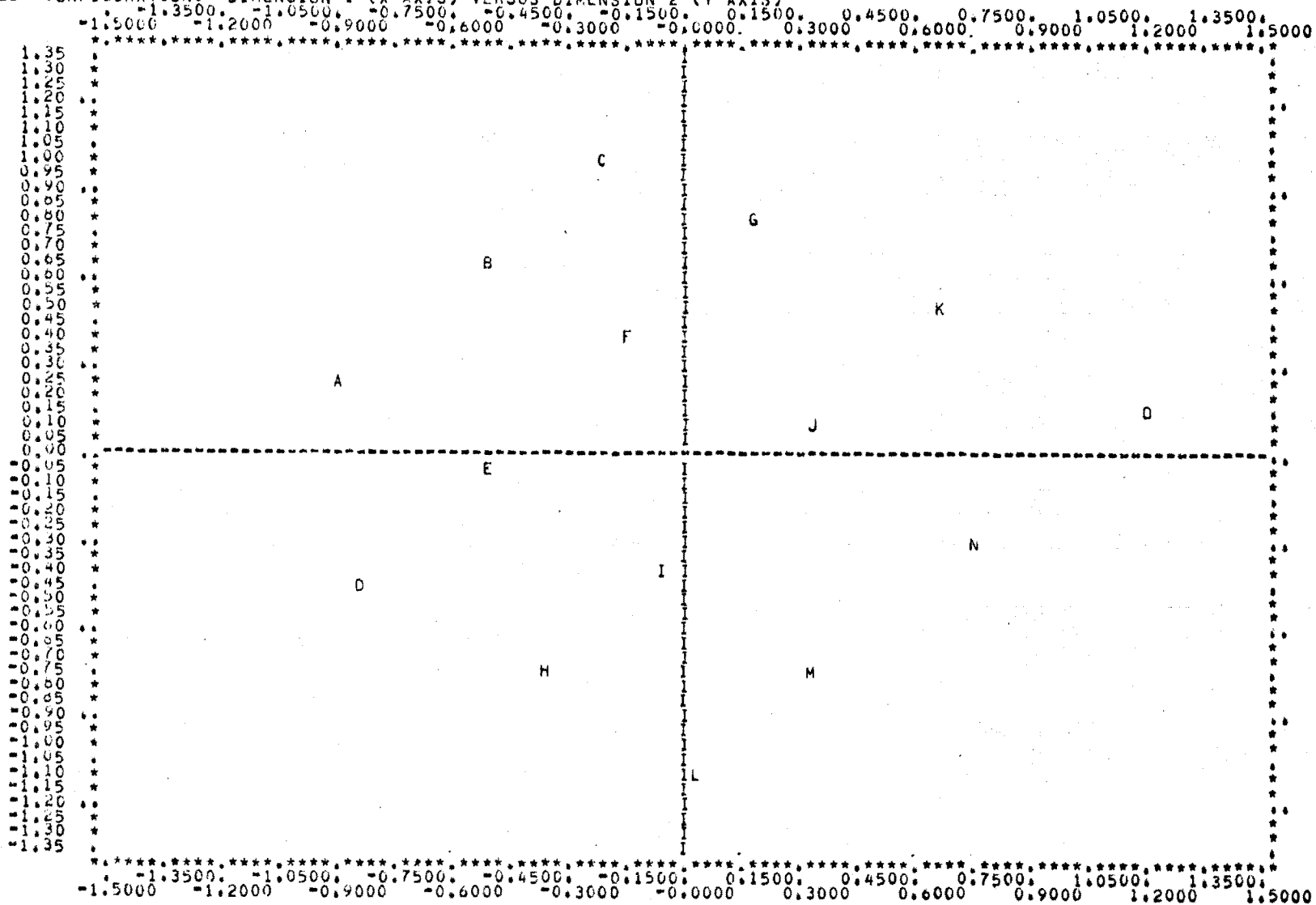


Figure 6.10: Monte Carlo Study 3 (see Table 6.5)

CONFIGURATION FROM MONTECARLO STUDYS, FROM EISLER CONFIGURATION, TO BE ROTATED TO THE EISLER CONFIG. IN TERMS OF POLARS.
ROTATED CONFIGURATION. DIMENSION 1 (X-AXIS) VERSUS DIMENSION 2 (Y-AXIS).

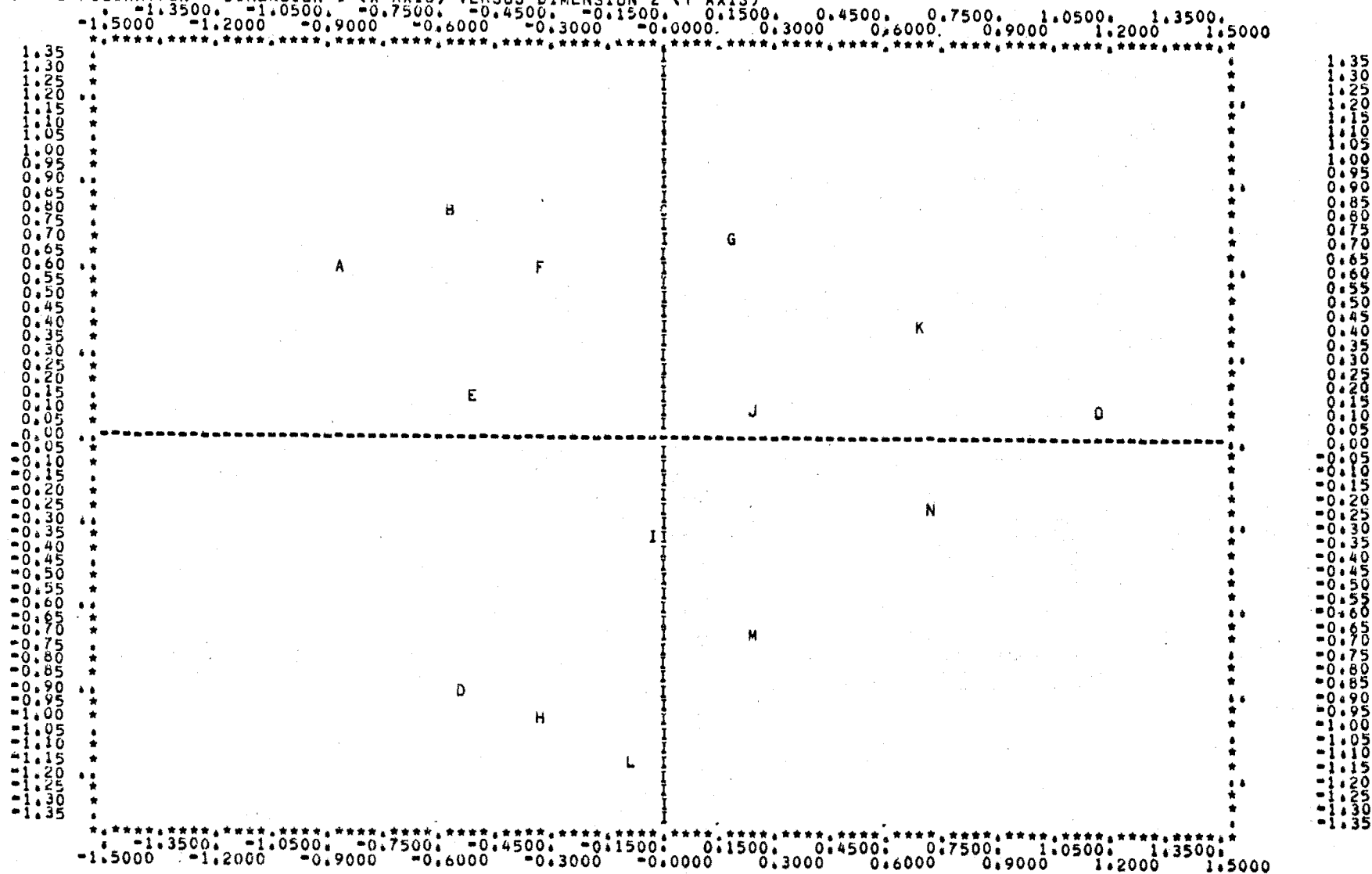


Figure 6.11: Monte Carlo Study 5 (see Table 6.5).

CONFIGURATION FROM MONTECARLO STUDY 8, FROM EISLER CONFIGURATION, TO BE ROTATED TO THE EISLER CONFIG. IN TERMS OF CARTESIANS.
 ROTATED CONFIGURATION: DIMENSION 1 (X-AXIS) VERSUS DIMENSION 2 (Y-AXIS)

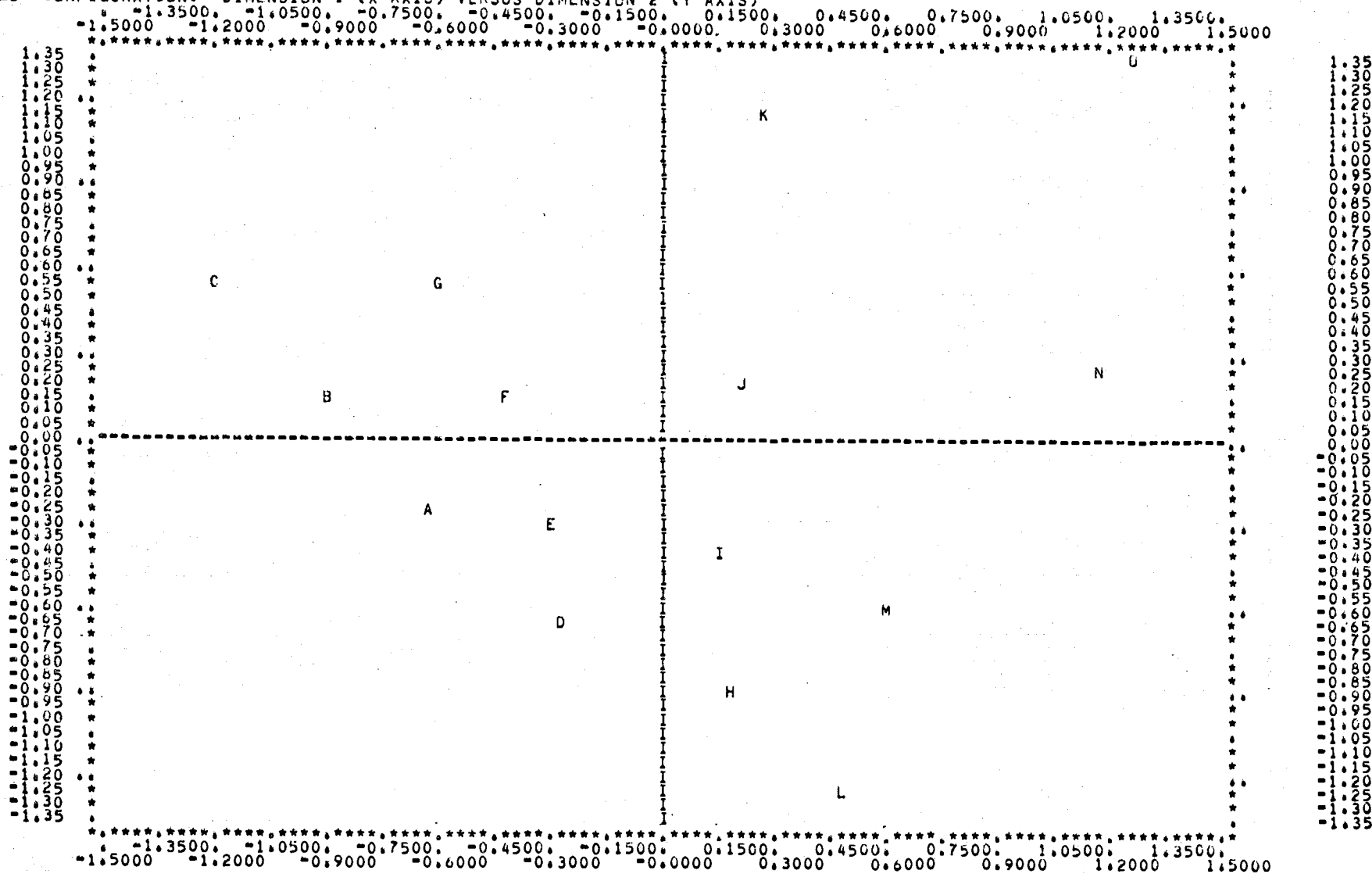
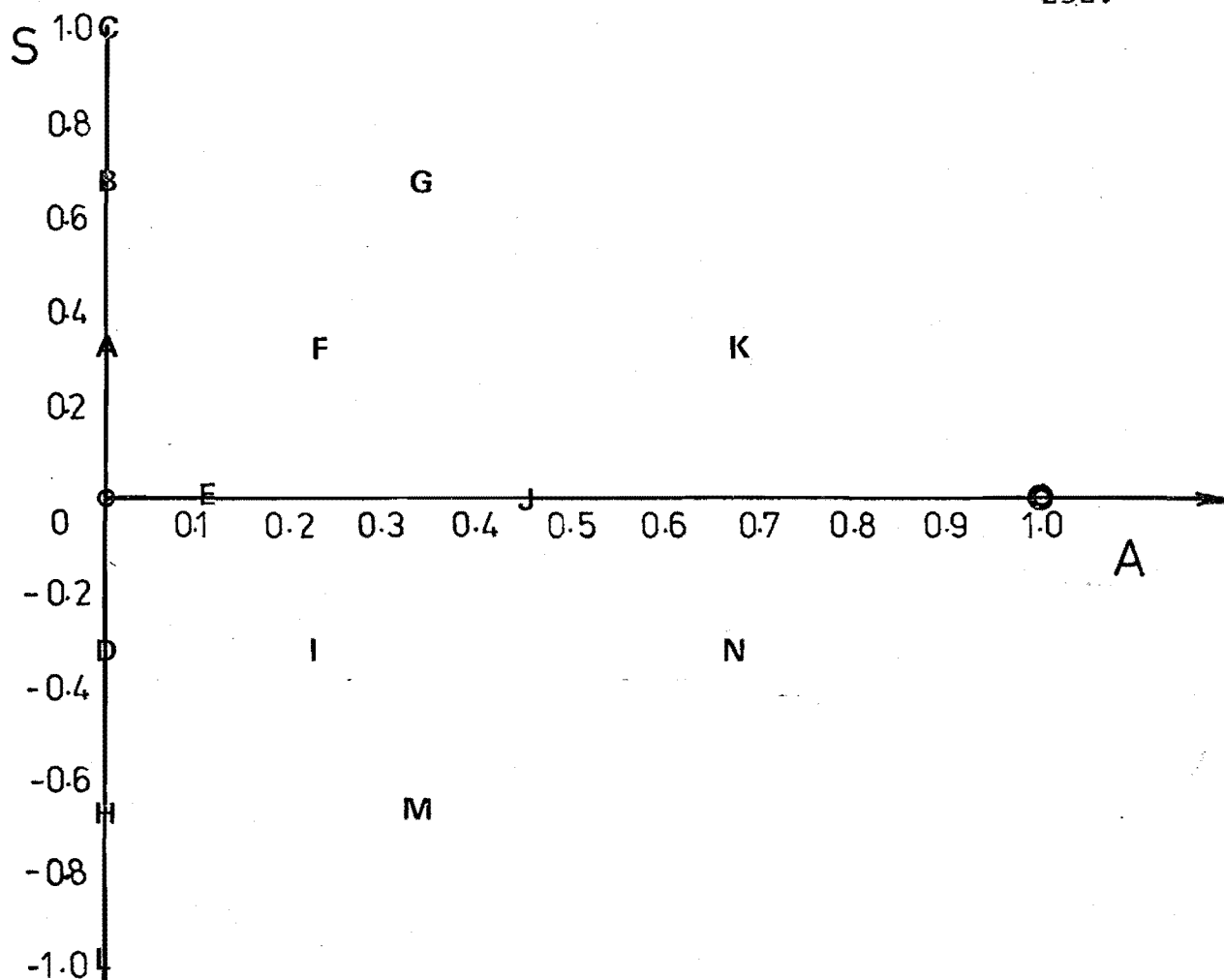
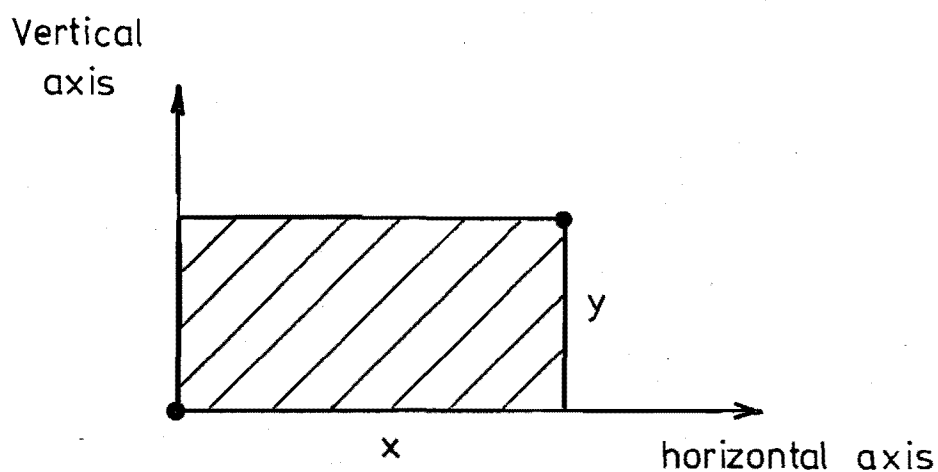


Figure 6.12: Monte Carlo Study 8 (see Table 6.5)



(a) Eisler Configuration in terms of Area ($=xy$) and "shape" ($=y-x$) of rectangle in (b).



(b) Rectangle to interpret (a). The two dots are two light-points of the experiment.

FIGURE 6.13 Eisler configuration in terms of area and "shape".

between the two is quite apparent. Figure 6.13(a) is the Eisler configuration expressed in terms of two parameters which are most simply described as area and "shape" of the rectangle shown in Figure 6.13(b). Thus if a point has horizontal and vertical components x and y respectively, the new parameters are $A=xy$ and $S=x-y$.

The scaling of study 7 is similar to study 8, but more distorted. Note that in all the above studies we have looked at only the first two dimensions of the rotated scaling.

Thus Eisler's algorithm appears to give valid results if the model describing the data is close to Eisler's model; but it can give quite misleading results (because they may be invalid but still interpretable) if the model is very different from Eisler's - and this includes the sometimes reasonable Minkowski models.

The results of the non-metric MDS (Minkowski) studies on POLYCON are summarised in Table 6.6, where the stress values tabulated are Kruskal's stress:

$$s^2 = \frac{\sum_i \sum_j (d_{ij} - \hat{d}_{ij})^2}{\sum_i \sum_j d_{ij}^2}$$

TABLE 6.6

Results of Monte Carlo studies of POLYCON

<u>Configuration</u>	<u>Dissimilarity model</u>	<u>Scaling model</u>	<u>Stress in</u>	
			2 Dims	3 Dims
P	D ₁₁	M ₁	.105	.081
P	D ₁₁	M ₂	.063 *	.027 §
P	D ₁₂	M ₁	.052	.035
P	D ₁₂	M ₂	.002	.000
P	D ₂₁	M ₁	.105	.092 *
P	D ₂₁	M ₂	.063 *	.027 §
P	D ₂₂	M ₁	.051	.041 +
P	D ₂₂	M ₂	.005	.001
SL	D ₁₁	M ₁	.044	.033 +
SL	D ₁₁	M ₂	.057 *	.036
SL	D ₁₂	M ₁	.034	.017
SL	D ₁₂	M ₂	.002	.000
SL	D ₂₁	M ₁	.043 *	.027
SL	D ₂₁	M ₂	.057 *	.036 §
SL	D ₂₂	M ₁	.049	.019
SL	D ₂₂	M ₂	.015	.004

For explanation of the P-SL notation, see Table 6.5.

Footnotes: * A minimum was found at this stress value; otherwise stress was very close to a minimum unless marked "+".

§ The third dimension has a significant weighting.

where the d_{ij} are the Minkowski distances of the derived configuration, and the \hat{d}_{ij} are the corresponding "disparities", monotone with the original data.

It will be seen from the table that the difference in stress from the addition of a third dimension is generally not enough to warrant consideration of a three-dimensional solution (and this is confirmed by Young's, 1968, index of metric determinacy, if in fact it is valid); in any case, as can be seen from the table, the derived configuration rarely had much weight on the third dimension. We therefore need consider only the two-dimensional solutions. It will also be noted that there is very little difference in stress between the corresponding $D_{1\beta}$ and $D_{2\beta}$ dissimilarity models; we thus consider only the $D_{1\beta}$ models in two dimensions.

When the configuration used to generate the dissimilarities is that expressed in terms of sines of elevations and vector lengths, the derived configuration generally is roughly faithful to the original, although there is definite distortion. A typical example (the lowest stress of such solutions) is given in Figure 6.14. This should be compared with Figure 6.5(d). Although the reproduction of the original configuration is by no means perfect, it is close enough so that it would be recognised in an empirical case. Note that,

EISLER CONFIGURATION: DISTANCES COMPUTED FROM CONFIGURATION EXPRESSED IN TERMS OF SINES OF ANGLE AND LENGTH
 DERIVED CONFIGURATION: DIMENSION 1 (X-AXIS) VERSUS DIMENSION 2 (Y-AXIS)

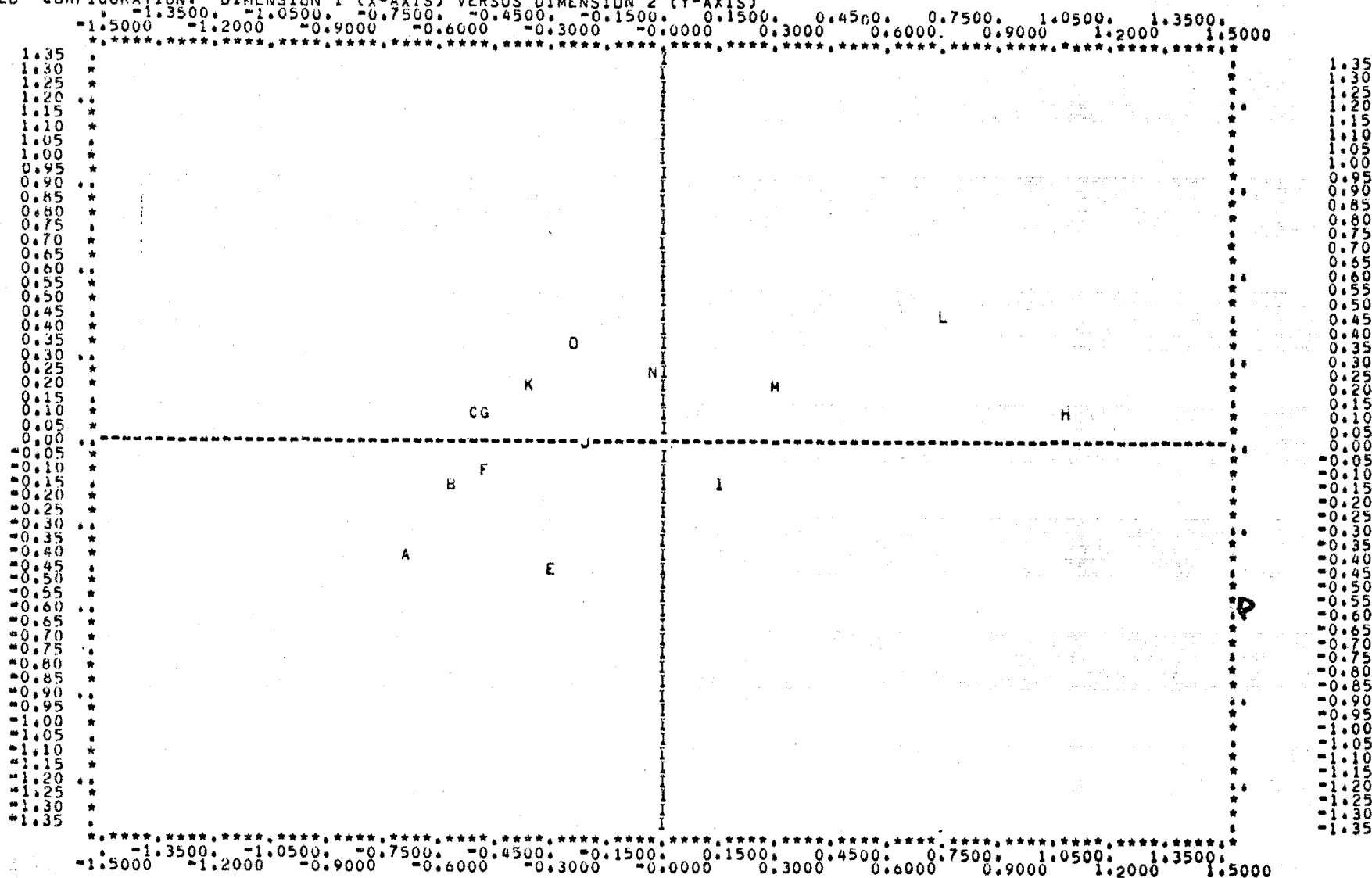


Figure 6.14: POLYCON Monte Carlo Study:
 Euclidean analysis of D_{12} dissimilarities. Stress = 0.2%

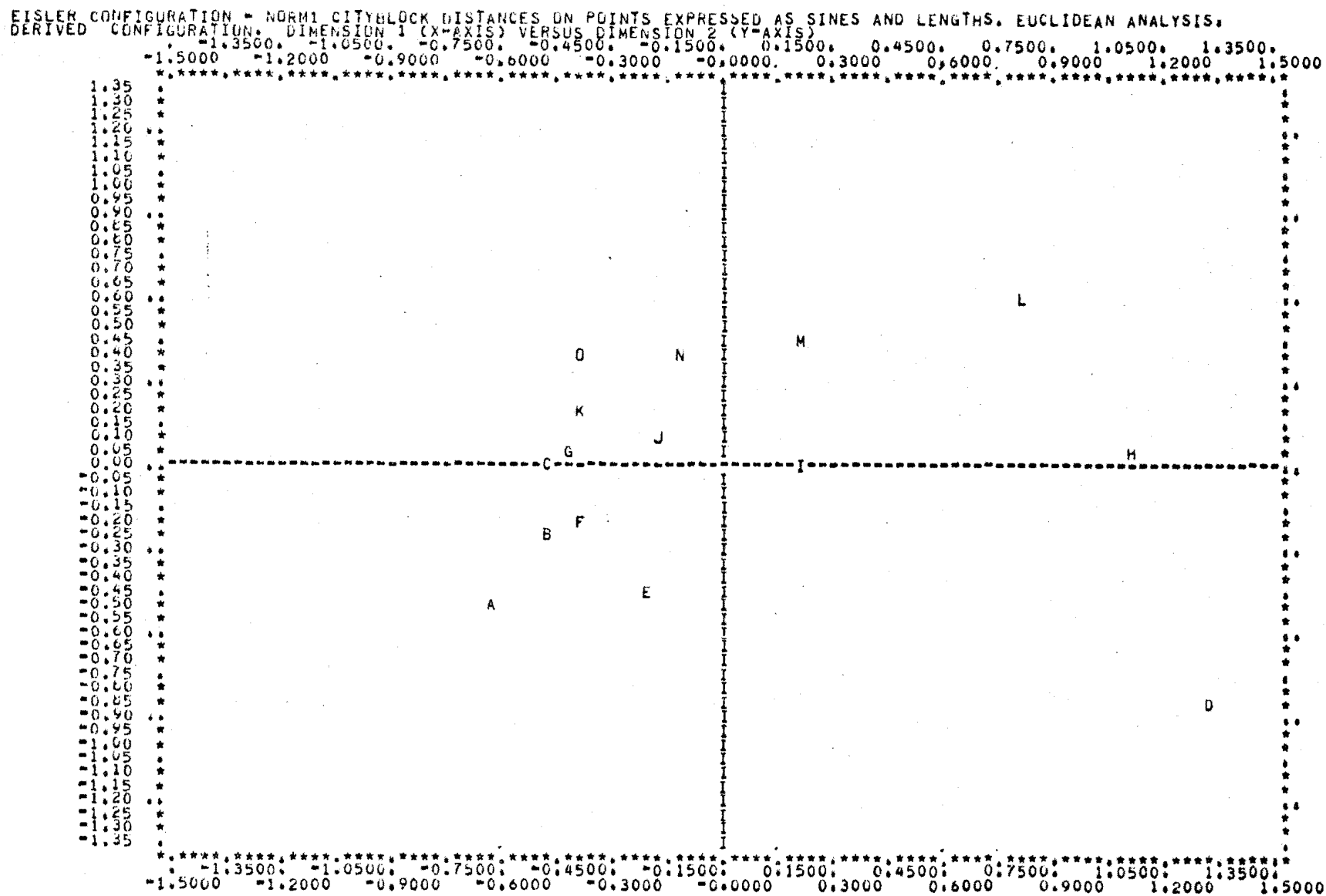


Figure 6.15: POLYCON Monte Carlo Study: Euclidean analysis of D_{11} dissimilarities, Stress = 5.7%

EISLER CONFIGURATION: NORMALISED EUCLIDEAN DISTANCES FROM PROJECTIONS ON AXES. EUCLIDEAN ANALYSIS.
 DERIVED CONFIGURATION: DIMENSION 1 (X-AXIS) VERSUS DIMENSION 2 (Y-AXIS)

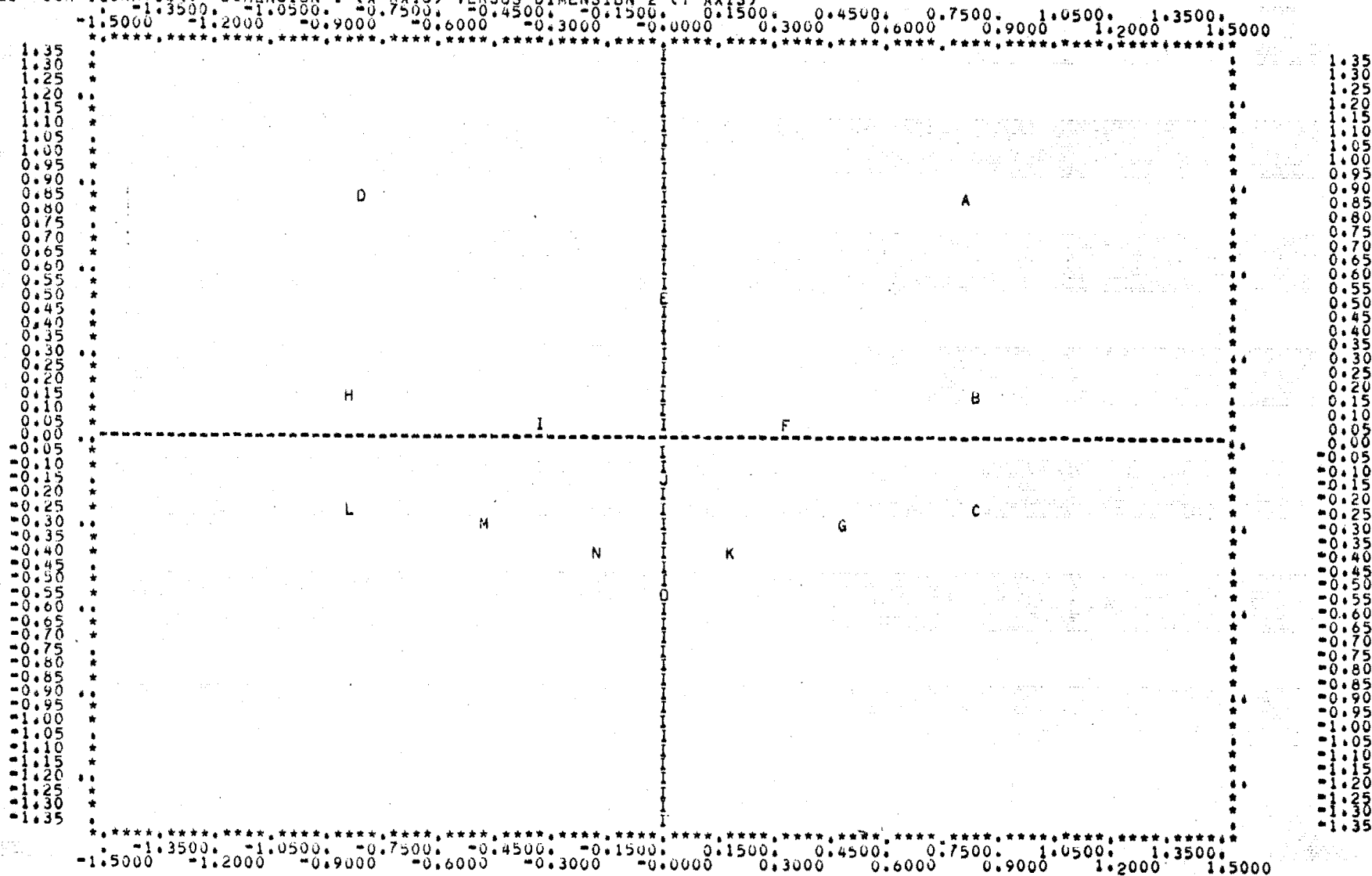


Figure 6.16: POLYCON Monte Carlo Study: Euclidean analysis of 12 dissimilarities. Stress = 0.2%

EISLER CONFIGURATION: NORMALISED CITYBLOCK DISTANCES FROM PROJECTIONS ON AXES. CITYBLOCK ANALYSIS.

DERIVED CONFIGURATION: DIMENSION 1 (X-AXIS) VERSUS DIMENSION 2 (Y-AXIS)

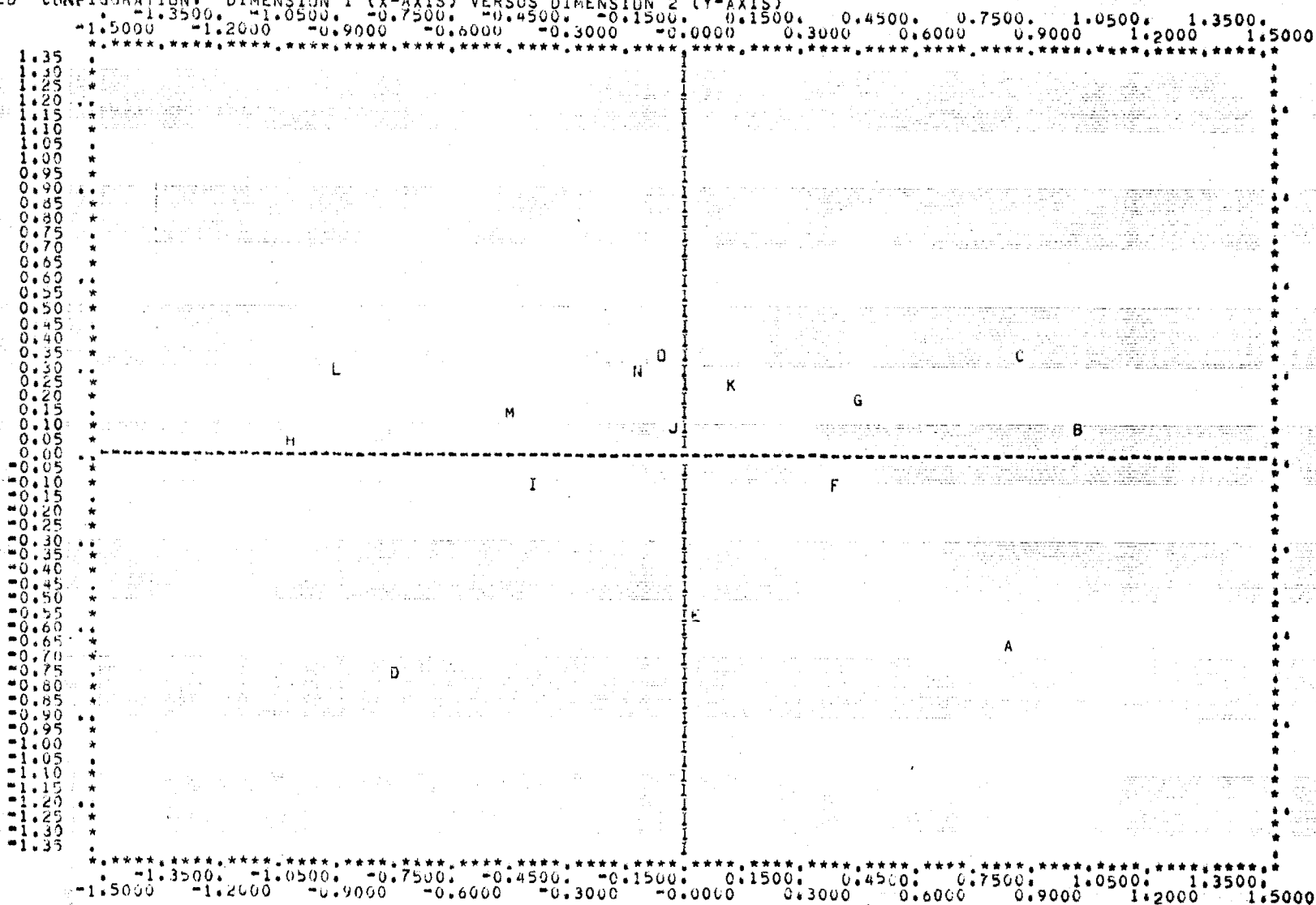


Figure 6.17: POLYCON Monte Carlo Study: City-block analysis of D₁₁ dissimilarities. Stress = 10.5%

although the dimensions of the solution are roughly correct, the model is not - even though it can give a very low stress (here, 0.2%; and in the other cases, not more than 5.7%: see Figure 6.15). Thus low stress and interpretable dimensions are not necessarily evidence for the Minkowski model.

When the original configuration is that expressed in terms of horizontal and vertical components, the stress is again low - not more than 10.5% (see Figure 6.17), but as small as 0.2% (compare the solutions found by Roskam, using Eisler's data: the minimum stress found there was 9%). But most significantly, the derived configuration is quite unlike the original configuration (a grid) but nonetheless highly interpretable as being in the space of angle of elevation versus vector length. The lowest stress (0.2%) solution is given in Figure 6.16 (to be compared with Figure 6.5(c)). The other such solutions are more distorted versions of this, but are still quite recognisable: again recognisable enough to lead to this incorrect interpretation in an empirical situation.

Thus Minkowski scaling, like Eisler's algorithm, leads to the generation of low stress, highly interpretable - but quite misleading - solutions if the model describing the data is similar to the $D_{1\beta}$ or $D_{2\beta}$ models. It appears to preserve dimensionality.

6.2.3 Review of Eisler's experiment

With these Monte Carlo results in mind, we can now review the analyses of Eisler's results carried out by Eisler and Roskam. There are two basic questions about this experiment that need answers. What type of model is being used - a normalised model (like Eisler's, or like $D_{1\beta}$ or $D_{2\beta}$), or a distance model (a Minkowski model)? And what dimensions are being used: projections on the two natural axes, or some function of the angle of elevation, and vector lengths?

The Monte Carlo results show clearly that, if a MDS approach is used to answer these two questions, they are not independent of one another. If the "real" configuration is in the space of sines and lengths, then, at least in terms of recovered dimensions, it makes little difference what the "true" (dis)similarity model is, or which model is used in the MDS algorithm. But if the "real" configuration is in the space of horizontal versus vertical, then these dimensions will be recovered only if the correct models are used. If the correct similarity model is a distance one and scaling is with a normalised model, the space will appear to be the area-shape one. If the correct similarity model is a normalised one and a distance model scaling is used then the apparent space will be an angle versus length one.

Since this last space is quite similar to the sine versus length space (especially considering that other distortions due to incorrect modelling can take place), the most plausible explanation of the discrepancy between the results of the two types of scalings carried out by Eisler and Roskam is that the subjective space is in fact the horizontal-vertical one, and the similarities conform best to a normalised type of model.

Thus Eisler's experiment may be taken as evidence for similarity models of the normalised type. Just which normalised model is the correct one will be decided only after a good deal more investigation; it may need the development of more flexible MDS algorithms, allowing a greater variety of (normalised) models, and requiring only (dis)similarity data input. It should also be remembered that not all so-called Content models will necessarily fall into the same category as the present "normalised" models: likely candidates for exclusion are Waern's multiplicative model (e.g. Waern, 1968a) and Hoijer's additive model (e.g. Hoijer, 1969a).

6.3 SOME ORDINAL DATA

In this section we shall very briefly consider - in terms of the present models - the few results from the ordinal approach to investigating similarity.

Tversky and Krantz (Tversky and Krantz, 1969, and Krantz and Tversky, 1975) and Wender (1971) have carried out similarity experiments in which the ordinal properties of various similarity models were tested (ref. Chapter 1). None of these studies use pooled data.

Tversky and Krantz (1969) used schematic faces in which the shape of the head, the eyes, and the shape of the mouth all could take two possible forms. They found that the results largely supported the ordinal properties of interdimensionally additive (and therefore decomposable) models - and so, in particular, Minkowski distances. Because of the simplicity (two-valued) and the nature of the dimensions in the experiment, it would seem plausible that they could be regarded as metathetic or qualitative by subjects. Thus we would indeed expect the Distance model to hold. Becker and Pipahl (1974) made a related observation, stating that the stimuli could be regarded as "nested" (which does not seem very credible). They suggested that the nested property was necessary for decomposability. This is exactly the opposite to the present theory; but it recognises that different types of continua may lead to different similarity results. Their result is due to the use of Restle's (dis)similarity models (see section 2.1.3.1) instead of the normalised ones we use (see equations 3.1 and 3.2).

Wender (1971) carried out a very similar experiment on rectangles; he found that the dimensions (which he assumed to be area and shape) interacted, so the similarities could not be Minkowskian. If the dimensions are in fact area and shape (defined as the ratio of height to width) then one (area) is clearly prothetic (see Stevens and Galanter, 1957) while the other would seem to be more metathetic (perhaps by analogy with the continuum of proportion: ref. op.cit.). Since metathetic continua are reversible (it makes equal sense here to take the ratio of width to height as that of height to width) Wender's results are in line with the present model for a mixture of prothetic and metathetic dimensions, which predicts an apparent interaction between dimensions (see the last equation in Chapter 3). The direction of the interaction varies over subjects in Wender's experiment (see especially his Table 2); this may correspond to the direction of the shape scale used by each subject. However some subjects show no interaction at all (see especially his Table 1); this could be partially ascribed to individual dimensional weighting, but Wender's results can only be described as equivocal: they reject the Distance model, but do not entirely support the present models. Becker and Pipahl explain Wender's results by erroneously claiming the dimensions of the space are not nested ones.

Most recently, Krantz and Tversky (1975) again investigated similarity between rectangles by ordinal methods. They found that neither the area-shape nor the width-height dimensional structures allowed dissimilarity to satisfy interdimensional additivity or intradimensional subtractivity. They rejected decomposability. They therefore had to reject the Minkowski distance models. Their more detailed finding was however that "an interval along one dimension appears longer the higher or more extreme the level of the orthogonal dimension" (op.cit., p.31). This is in accord with an observation by Fenker (1972), but is exactly the opposite tendency to that predicted by our mixed model when "the orthogonal dimension" is the prothetic one. Krantz and Tversky do remark that both tendencies can occur: which one occurs depends on the stimulus set. This experiment therefore apparently provides a counterexample to any of the models of this thesis.

6.4 CONCLUSIONS

The data surveyed in this chapter is of course only a very small sample of the huge body of similarity data available. It does however give quite strong, though qualified, support to the models of this thesis. In section 6.1.1 we showed that the present models gave very good account of data on four quite different stimulus sets, reducing the dimensionality and contributing to the understanding of the revealed

dimensions.

In the last section, the models were able to explain at least some of the results from ordinal investigations of similarity. This area is an important one in that it is one of the few in which the mathematics of the modelling involved has been rigorously investigated (by Beals, Krantz and Tversky).

Thus the models are seen to be reasonable ones, worthy of deeper investigation. An important warning is sounded by the experiments in the last section, however, on pooling data. As Wender's experiment shows in particular, although interactions between dimensions definitely occur, those interactions can vary in direction between subjects. Pooling of such subjects would "cancel out" such interactions, perhaps leaving no apparent interaction, but at least rendering any that is then present quite unrepresentative of the true state of affairs. Therefore great care should be taken when pooling data from multidimensional spaces where the direction of a dimension is ambiguous and the dimensions are either all prothetic, or mixed metathetic and prothetic.

The investigation of MDS algorithms by Monte Carlo methods shows quite dramatically that, even when problems of local minima have been overcome, the model

used in the scaling can significantly affect the validity of the MDS solution. An incorrect scaling model can lead to a low stress, interpretable solution which is nonetheless quite invalid.

MDS algorithms also do not appear to be very sensitive to different models within a certain category. For example, all Minkowski models tend to give very similar solutions, as do all models of a normalised type (including most of the usual Content models: see Table 2.1). Between types of model there may be quite radical differences in number and interpretation of the dimensions of the solution; within types, the main difference is in solution goodness of fit (stress), though dimensionality may also be affected.

From our investigations it appears that both these model types may be valid, but in differing contexts: the Minkowski models generally in "metathetic" situations, the normalised models generally in "prothetic" situations. Further - more detailed - research is needed to identify precisely which model or models are correct within each model type.

CHAPTER 7

DISCUSSION AND CONCLUSIONS

In Chapter 1, the main aim of this dissertation was stated as being to investigate

- a) the adequacy of present distance functions for modelling psychological similarity, either in the form of direct judgments, or as implicit in other behaviour; and
- b) the adequacy of present distance functions for use in psychophysical scaling.

"Present distance functions" were shown to be distinguished by three "gross" properties:

$$D(x,y) \leq D(x,z) + D(z,y) \quad \forall x,y,z \in \mathbb{R}^n \quad (7.1)$$

$$D(\alpha x, \alpha y) = |\alpha| D(x,y), \quad \forall \alpha \in \mathbb{R}, x,y \in \mathbb{R}^n \quad (7.2)$$

$$\text{and } D(x+a, y+a) = D(x,y) \quad \forall a, x,y \in \mathbb{R}^n \quad (7.3)$$

It was suggested that the place to look for a more reasonable distance function, at least on prothetic continua, was among the Content models of similarity; these have the contrasting gross features

$$D(\alpha x, \alpha y) = D(x,y) \quad \forall \alpha \in \mathbb{R}, x,y \in \mathbb{R}^n \quad (7.4)$$

$$D(x+a, y+a) < D(x,y) \quad x,y,a \in \mathbb{R}^{n+} \quad (7.5)$$

which are mathematically incompatible with (7.2) and (7.3). Working from a set-theoretic approach to the Content models, two extreme cases of distance functions

were discovered. For the prothetic (quantitative) case, the normalised types of model, $D_{1\beta}$ and $D_{2\beta}$ seem plausible; for the metathetic (qualitative) case, the Minkowski distances, M_β , seem plausible.

We thus have a quite new type of distance function for quantitative dimensions. It gives to the space an immoveable origin and, except for a specific case, fixed axes. Its effect on the space is reminiscent of a perspective transformation: pairs of stimuli having the same difference on a dimension will get closer together as they get further from the origin (like railway tracks in the distance). But equal proportionality between stimuli maintains their distances: two stimuli are the same distance apart at any distance from the origin as long as each respective pair of components is always in the same ratio. Perhaps the greatest practical importance of these properties is that as a result, the $D_{1\beta}$ and $D_{2\beta}$ functions cannot be monotone with the M_β functions. Thus even non-metric M_β -based MDS of prothetic spaces would be invalid.

If it can be shown that the normalised models are good ones, then much similarity modelling will have to be rethought. We shall in this chapter briefly review some of the strengths and weaknesses of our approach to similarity modelling to give a clearer picture of the point to which this thesis has brought the psychology of similarity.

7.1 THE MODELLING ASSUMPTIONS

The theoretical development of our models is almost entirely based on the set theoretic view of psychology. Restle gave this viewpoint respectability with his persuasive analysis of aspects of judgment and choice; but doubt still quite rightly remains as to its validity. Luce and Galanter's comments on this matter have already been quoted in Chapter 2 (section 2.1.3); the main ground for doubt is the question of whether explaining psychological phenomena in terms of sets really "explains" anything. At the present there is no real answer to this - no basis (the most likely seems to be a physiological one) for the sets and their elements has even been formally proposed. So the only defence is that the approach seems to work. It can only be hoped that a deeper justification may flow from increased interest and data generated by success. One might point out though, that the set theoretic approach certainly has far more intuitive attraction than the "vector" approach widely used by Content model workers.

Once a set theoretic approach is accepted however, there are two assumptions of Restle's which the development of Chapter 3 brought into question. First was the definition of measure, m , which usually assumes simply that the measures of disjoint sets add to give the measure of the union. It may be useful to make additional or different assumptions as to particular

cases of the union of non-disjoint sets, and as to the union of disjoint sets. One suggestion already made was

$$\underline{m}(X \cup Y) = (\underline{m}(X)^\beta + \underline{m}(Y)^\beta)^{1/\beta}, X \cap Y = \phi, \beta > 1.$$

Measures taking negative values may also be worth developing.

The second assumption of Restle's was that the only arrays worth considering were the metathetic arrays (of which prothetic arrays are a special case) because they preserved betweenness, and additivity of the distance function Restle defined:

$$d(X, Y) = \underline{m}(X \Delta Y).$$

Since we are now using a quite different distance function (see Chapter 3), it seems likely that other types of array may be worth consideration. It is possible that a study of other such arrays, and also of particular cases of arrays midway between prothetic and equal-measure metathetic, would be fruitful in drawing further distinctions between subjective continua. It may also be worthwhile reconsidering our assumption that different dimensions are disjoint (see assumptions P.(c) and M.(e) of Chapter 3): a certain degree of communality or interaction between dimensions may be realistic in some situations.

While on the subject of types of continua, it should be noted that the suggestions made in early chapters, that the "equal-intensity" concept of the Content theorists might correspond to our equal-measure continua, have been shown to be largely groundless. This does not however close the door entirely on the idea that some of the "equal-intensity" continua (which are usually mapped into at least two dimensions) may reduce to unidimensional continua of a special type which may be worth further investigation in their own right. What may be happening is that the present models tap a metathetic quality which is due primarily to physiological factors, but which does not necessarily coincide with the intuitive notion of "qualitative" (metathetic) tapped by "equal-intensity" continua.

Perhaps the most outrageous step we made was that of generalising the set-theory derived D_{11} and D_{21} to the functions $D_{1\beta}$ and $D_{2\beta}$ for a wide range of β . It is no more outrageous, though, than the similar generalisation made from Euclidean (M_2) and City-Block (M_1) distances to any Minkowski distance (M_β). Again, it will stand or fall on its empirical support.

7.2 THE MODELS

Some of the more obvious and interesting properties of the $D_{1\beta}$ and $D_{2\beta}$ models were spelt out in Chapter 4. They have most of the properties thought desirable for the prothetic case in Chapter 1. They have other interesting properties such as non-convex isosimilarity contours, and many breaches of the triangle inequality. The one dimensional case (either embedded in a multi-dimensional space, or alone) is particularly interesting with the triangle inequality holding strictly so that the commonly assumed property of additivity on straight lines never holds, at least on lines through the origin.

Of $D_{1\beta}$ and $D_{2\beta}$, $D_{1\beta}$ seems the better behaved mathematically. $D_{2\beta}$ violates the triangle inequality more significantly than $D_{1\beta}$ and additionally, if negative components are permitted (a debatable point), is not bounded by unity, and is undefined for points exactly on opposite sides of the origin. On the other hand, D_{22} is very interesting because its isosimilarity contours are off-centre circles, and so it may be indistinguishable from Euclidean distances under certain tests. On the whole though, these two normalised models are not radically different.

Of almost equal interest to the dissimilarity models themselves are the two suggested similarity gradients

$$F_1(\alpha) = 1 - \alpha \quad 0 \leq \alpha \leq 1 \quad (7.6)$$

$$F_2(\alpha) = \frac{1 - \alpha}{1 + \alpha} \quad 0 \leq \alpha \leq 1 \quad (7.7)$$

F_1 has been widely suggested previously, but few cogent reasons have been given for adopting it; F_2 seems worth greater consideration, especially considering the empirical findings of Chapter 6. It is interesting (though it probably has no significance) that both F_1 and F_2 are idempotent:

$$F_1(F_1(\alpha)) = F_2(F_2(\alpha)) = \alpha.$$

Both gradients apply for both prothetic and metathetic situations according to the theory (although Chapter 6 suggests otherwise); this implies restrictions on the scale used in the metathetic (Minkowski) case, since otherwise similarities will become negative.

One problem that the set theoretic approach is not helpful on is that of the status of the scales used in the models: are they subjective (psychological) or objective (physical)? It would seem more reasonable that the models are in subjective space; this is supported by the data in Chapter 6. But at least one of the models (Gregson's) that are special cases of the present functions, was developed largely for objective

values (though see Gregson, 1975, p. 235-238), solving only for dimensional weights or other similar parameters. And some of the implications of the models in other areas of psychology (see next section) would not be true if the model is in subjective space. However, if subjective scales have any real validity, in terms of truly reflecting the perceptions of an individual, then it would be amazing if they are not used in similarity judgments. Of course it could be the case that both the unidimensional magnitude estimations and the similarity judgments are summarised in the one model. The problem of on what basis unidimensional judgments are made (if at all) within the similarity judgment, compared to the corresponding judgment in isolation, is theoretically an important, interesting and perhaps unsolvable one. In practice there may be not enough difference to worry about. But it seems safest to assume, particularly with regard to empirical evidence presented here, that the models are in the psychological space defined by direct magnitude estimations.

A related question is the scale type: are the unidimensional scales ordinal, interval or ratio scales? Or something else? The form of the $D_{1\beta}$ and $D_{2\beta}$ models dictates that the prothetic scales must be at least ratio scales (the models themselves are on absolute scales). The form of the Minkowski models

dictates at least an interval scale for the metathetic continua, with the models themselves (as is often forgotten in MDS applications) on ratio scales.

Many of the proofs of the properties of the $D_{1\beta}$ and $D_{2\beta}$ models presented here are weak, and occasionally even nonexistent. The investigation here is no more than an attempt to get an initial "feeling for" these models; if they prove useful, then more rigorous investigation should be carried out. A different approach to the proofs may be by using directly the set theoretic foundation to prove many of the properties in a more general context. Tversky (1972) has given examples of such proofs in a different context.

7.3 APPLICATIONS TO OTHER AREAS

An attempt was made to apply the similarity models to other areas of psychology, particularly category scaling. On the whole it was remarkably successful: successful enough to make a similarity interpretation worth more detailed study. Its main failing is that it does not supply a rational means for taking into account the variety of context effects (experimental conditions) which are known to affect category judgments. Two parameters seemed sufficient to take many of these effects into account however. Of particular interest was the non-logarithmic model.

The most notable achievement of the present models was however to "explain" why the relation between magnitude estimation and category scales is different for metathetic and prothetic continua. Thus the Stevens and Galanter and the Restle concepts of metatheticness and protheticness are neatly tied together. This result adds considerable weight to the set theoretic and similarity views of scaling.

The short foray into discrimination (Ekman's Law) seemed quite successful in view of the result on colour vision in Chapter 6. One might suggest from this that our "similarity" models are in fact ones of discrimination (they are certainly closely allied to several suggested for discrimination). At the least, there is a close relationship between the two behaviour patterns.

These applications should not be taken too seriously at this point: the similarity models will not fall if they do not succeed. But they certainly will be strengthened if the apparent connections we have indicated can be shown to be more than merely fortuitous. Similarity was first studied because it seemed basic to many types of behaviour. There seems little reason for continuing to study it (no matter how well similarity judgments per se can be predicted) if it cannot be fed back to make predictions in the areas from which it first arose.

7.4 EMPIRICAL EVIDENCE

Very little direct evidence is presented here in support of the multidimensional models. This is mainly due to the lack of suitable MDS algorithms, which seem the simplest way of testing them directly. What evidence is presented, though, gives firm support to the unidimensional models and their interpretation of metathetic and prothetic continua. The multidimensional example shows that a normalised type of model is the most reasonable one, but to decide just which model it is requires further work. The ordinal results gave only qualified support to the models. They do show that the models have a definite part to play that cannot be played simply by the usual distance models.

On the whole we can conclude the following. Evidence presented here suggests there is enough support for the models to warrant their further investigation. But there are contexts in which they do not apply (one example of these is in the experiment of Krantz and Tversky, 1975; others would include the type of similarity effect - "holes in the space" - found by Goldmeier, 1972). The models apparently take insufficient account of non-uniformities in the space. They are plausible, but not entirely accurate models, at least as so far developed.

7.5 MULTIDIMENSIONAL SCALING

The model on prothetic continua has been shown to be at the least a plausible one: many of its most important features will be present in any other such similarity model. Therefore the findings in MDS, in Chapter 6, have implications that are not tied merely to the present models. It was shown that by using the wrong model in the MDS algorithm one could get highly interpretable low stress solutions which were quite invalid descriptions of the process generating the data. This implies that many of the scalings done in the past may be quite incorrect and misleading.

The following seems to be the position as it now stands. There are two general types of model: normalised types (including the present prothetic models and most Content models) and distance types. The two are quite incompatible as far as, firstly, MDS is concerned, and, secondly, the type of stimulus judged is concerned. Each type of model applies to one type of stimulus (prothetic or equal-measure metathetic) and scaling of one type of model by the other type leads to quite invalid and highly misleading results. (The position is of course further complicated by the existence of models midway between "distance" and "normalised").

Therefore, before any MDS is carried out, it should be determined, from the type of stimuli, which type of model is the more likely to hold. Only when this is decided should a suitable scaling be performed. Stress (goodness of fit) values have a valid meaning only within types of model - not between.

7.6 GENERAL CONCLUSIONS

What this thesis has done is to delineate, using the familiar prothetic-metathetic distinction between types of continua, just where the various models of similarity should hold true. This delineation has been given a *prima facie* validation in unidimensional psychophysics and the psychophysics of similarity.

Normalised types of distance functions, which are not necessarily metrics, probably of the form of our $D_{1\beta}$ or $D_{2\beta}$ functions, hold in the purely prothetic cases. Thus the "Content" type of model, the main examples of which are given in Table 2.1, and which are of the normalised type, should be applied only to the prothetic case.

Distance models hold in the purely metathetic "equal-measure" cases. This means that the Minkowski models should be used for modelling similarities between only purely qualitative (equal-measure metathetic) stimuli.

In addition to this delineation, we have shown that great care must be taken in applying the various models in MDS algorithms. That is, not only are the two types of model quite separate psychologically, but they are also quite different mathematically. So mixing models will give quite invalid results both psychologically and mathematically.

The mathematical difference between the two types of model is perhaps summed up in the fact that they are not monotonically related to each other. At its simplest, we have shown that there are psychologically very reasonable distance functions for which scaling with even non-metric MDS using Minkowski distances gives quite invalid results. That is, the flexibility given to MDS by allowing any monotonic transformation of similarity data (i.e., non-metric MDS) does not allow one to ignore the dissimilarity model which is the basis of the algorithm.

The main aims of this thesis, as set out in Chapter 1 and at the beginning of this chapter have therefore been achieved. A possible rational basis has been laid for the further development of similarity modelling and MDS.

7.7 SUGGESTIONS FOR FURTHER STUDY

The framework erected in this thesis is no more than that: it needs testing to see if it will bear the weight of empirical evidence. And if it is found to be strong enough, it needs to be filled in and its weak points strengthened. It then has to be used in new applications.

The testing of the model will initially have to be mainly on a basis of comparing the normalised versus the distance model types, to see which (if either) fits better in a given context. This can be done either directly through similarity estimates and model fitting, or indirectly, by testing for some of the distinctive "gross features" of the types of models. The "gross features" given in Chapter 1 would be obvious ones to test for initially; topological tests like that of Zagorski (1974) would also be worth developing. A most important point here will be to find exactly what "prothetic" and "equal-measure metathetic" mean in reality. To do this it will be necessary to (a) find some way of measuring "protheticness" (for example, in terms of discrimination variability - see Eisler, 1963) and (b) to specify some of the dissimilarity models that correspond to cases of continua of varying protheticness. It will also be necessary to develop the multidimensional models on dimensions of mixed types.

The more detailed testing of the models would include determining which (if either) of the two similarity gradients F_1 and F_2 are correct: again, they may be valid in different contexts, as evidence in Chapter 6 suggests. The development of a (perhaps non-metric) MDS algorithm for the normalised models may be useful both for applications of the models and for detailed testing of the $D_{1\beta}$ and $D_{2\beta}$ distance functions.

If this testing shows that the models are useful then the process of filling in holes and overcoming weaknesses would be worthwhile. The models could be investigated in non-dimensional (see Cunningham and Shepard, 1973; Boorman and Arabie, 1972; or Johnson, 1967) and discrete spaces: new (non-metathetic) set-theoretic arrays could be investigated. It may also be useful to attempt an axiomatisation of the normalised distance functions parallel to that of Beals, Krantz and Tversky. The significance of the β parameter could be investigated as has been done for the Minkowski case (e.g. Hyman and Well, 1967, 1968).

The models have many weaknesses. Goldmeier (1972, pp. 125-126) raises the problem of discontinuities in the similarity space; many context effects (see, for example, Goldmeier, op.cit., Fenker, 1972, and Torgerson, 1965) have not been accounted for. Consideration of the choice of "relevant dimensions"

and effect of "core", which we have ignored, may throw light on this. Individual differences between subjects need closer study in terms of these models - although the implicit or explicit (equation 3.29) dimensional weighting of our models may at least partially account for these. Closely related to this is the problem, brought up in Chapter 6, of choice by subjects of which dimensions to use.

Finally, it is worth investigating the similarity view of category scaling to see whether a more precise model (taking known context effects into account) is possible. The relationship between discrimination and similarity, as suggested in the similarity investigation of Ekman's Law, and supported in our analysis of the third experiment from Ekman (1965) also needs further study.

REFERENCES

- ADAMS, E.W., AND FAGOT, R.F. On the theory of biased bisection operations and their inverses. J. Math. Psychol., 1975, 12, 35-52.
- ARABIE, P. Concerning Monte Carlo evaluations of nonmetric multidimensional scaling algorithms. Psychometrika, 1973, 38, 607-608.
- ARABIE, P., AND BOORMAN, S.A. Multidimensional scaling of measures of distances between partitions. J. Math. Psychol., 1973, 10, 148-203.
- ARABIE, P., AND SHEPARD, R.N. Representation of similarities as additive combinations of discrete overlapping properties. Presented to the sixth Annual Math. Psychol. Meeting, Montreal, 1973.
- ARMSTRONG, J.S. Derivation of a theory by means of a factor analysis, or Tom Swift and his Electric Factor Analysis Machine. Amer. Statistician, 1967, December, 17-21.
- ATKINSON, R.C., BOWERS, G.H., AND CROTHERS, E.J. An introduction to mathematical learning theory. New York: Wiley, 1965.
- ATTNEAVE, F. Dimensions of similarity. Amer. J. Psychol., 1950, 63, 516-556.

- BÄCKSTRÖM, L., AND GOUDE, G. An empirical examination of a quantitative model for similarity experience. Rep. Dept. Psychol., Univ. Uppsala, 1966, No. 40.
- BEALS, R., AND KRANTZ, D.H. Metrics and geodesics induced by order relations. Mathematische Zeitschrift, 1967, 101, 285-298.
- BEALS, R., KRANTZ, D.H., AND TVERSKY, A. Foundations of multidimensional scaling. Psychol. Rev., 1968, 75, 127-142.
- BECHTEL, G.G. Comparative scaling of unidimensional discrimination and similarity data. Psychometrika, 1966, 31, 75-84.
- BECKENBACH, E., AND BELLMAN, R. An introduction to inequalities. New York: Random House, 1961.
- BECKENBACH, E., AND BELLMAN, R. Inequalities. Berlin, Heidelberg: Springer-Verlag, 1971.
- BECKER, J., AND PIPAHL, W. The additive difference model and a metric and an ordering on sets. Psychometrika, 1974, 39, 361-362.
- BINDRA, D., DONDERI, D.C., AND NISHISATO, S. Decision latencies of 'same' and 'different' judgments. Perception and Psychophysics, 1968, 3, 121-130.
- BLANK, A.A. The Luneburg theory of binocular visual space. J. Opt. Soc. Amer., 1953, 43, 717-727.

- BLANK, A.A. Axiomatics of binocular vision. The foundations of metric geometry in relation to space perception. J. Opt. Soc. Amer., 1958, 48, 328-334.
- BLOCK, J. Amendment to Cohen's cautions regarding the measurement of profile similarity. Psychol. Bull., 1970, 73, 307-308.
- BOORMAN, S.A., AND ARABIE, P. Structural measures and the method of sorting. In Shepard, R.N., Romney, A.K., and Nerlove, S.B. (Eds.) Multidimensional scaling: Theory and applications in the behavioural sciences. New York: Seminar Press, 1972, Vol. I.
- BORING, E.G. Gustav Theodor Fechner. In Newman, J.R. (Ed.) The world of mathematics. London: Allen and Unwin, 1960, Vol. II.
- BRAY, J.R., AND CURTIS, J.T. An ordination of the upland forest communities of Southern Wisconsin. Ecol. Monogr., 1957, 27, 325-349.
- BUSEMANN, H. The Geometry of Geodesics. New York: Academic Press, 1955.
- BUSH, R.R., AND MOSTELLER, F. A model for stimulus generalisation and discrimination. Psychol. Rev., 1951, 58, 413-423.
- CANTER, R.R., AND HIRSCH, J. An experimental comparison of several psychological scales of weight. Amer. J. Psychol., 1955, 68, 545-649.

- CARGO, D. A multidimensional scaling of similarity estimations. Unpublished project report, Psychology Department, University of Canterbury, 1973.
- CARROLL, J.D. Individual differences and multidimensional scaling. In Shepard, R.N., Romney, A.K., and Nerlove, S.B. (Eds.) Multidimensional scaling: Theory and applications in the behavioural sciences. New York: Seminar Press, 1972, Vol. 1.
- COHEN, J. r_c : a profile similarity coefficient invariant over variable reflection. Psychol. Bull., 1969, 71, 281-284.
- COOMBS, C.H. A theory of data. New York: Wiley, 1964.
- COOPER, L.G. A multivariate investigation of preferences. Multivariate Behav. Res., 1973, 8, 253-272.
- CUNNINGHAM, J.P., AND SHEPARD, R.N. Monotone mapping of similarities into a general metric space. Presented to the sixth Annual Math. Psychol. Meeting, Montreal, 1973.
- DERKS, P.L. Visual recognition of similarity and identity. J. Exp. Psychol., 1972, 95, 237-239.
- EGGLESTON, H.G. Convexity. Cambridge: Cambridge University Press, 1958. (Cambridge Tract in Mathematics and Mathematical Physics, No. 47).

- EISLER, H. Similarity in the continuum of heaviness with some methodological and theoretical considerations. Scand. J. Psychol., 1960, 1, 69-81.
- EISLER, H. On the problem of category scales in psychophysics. Scand. J. Psychol., 1962, 3, 81-87.
- EISLER, H. How prothetic is the continuum of smell? Scand J. Psychol., 1963, 4, 29-32.
- EISLER, H. A choice model for paired comparison data based on imperfectly nested sets. Psychometrika, 1964, 29, 363-370.
- EISLER, H. Multidimensional similarity: An experimental comparison between vector, distance, and set theoretic models. Unpubl. ms., 1967.
- EISLER, H., AND EKMAN, G. A mechanism of subjective similarity. Acta Psychologica, 1959, 16, 1-10.
- EKEHAMMAR, B. A comparative study of some multidimensional vector models for subjective similarity. Scand. J. Psychol., 1972, 13, 190-197 (a).
- EKEHAMMAR, B. Multidimensional scaling according to different vector models for subjective similarity. Acta Psychologica, 1972, 36, 79-84 (b).
- EKMAN, G. Some aspects of psychophysical research. In Rosenblith, W.A. (Ed.) Sensory communication. New York: Wiley, 1961.

EKMAN, G. A direct method for multidimensional ratio scaling. Psychometrika, 1963, 28, 33-41.

EKMAN, G. Two methods for the analysis of perceptual dimensionality. Percept. Mot. Skills, 1965, 20, 557-572.

EKMAN, G., ENGEN, T., KUNNAPAS, T., AND LINDMAN, R.
A quantitative principle of qualitative similarity.
Rep. Psychol. Lab., Univ. Stockholm, 1963, No. 152,

EKMAN, G., ENGEN, T., KUNNAPAS, T., AND LINDMAN, R.
A quantitative principle of qualitative similarity.
J. Exp. Psychol., 1964, 68, 530-536.

EKMAN, G., GOUDE, G., AND WAERN, Y. Subjective similarity in two perceptual continua. J. Exp. Psychol., 1961, 61, 222-227.

EKMAN, G., AND LINDMAN, R. Multidimensional ratio scaling and multidimensional similarity. Rep. Psychol. Lab., Univ. Stockholm, 1961, No. 103.

EKMAN, G., AND SJÖBERG, L. Scaling. Annu. Rev. Psychol., 1965, 16, 451-474.

FENKER, R.M. Tensor analysis and multidimensional scaling: The influence of class structure on similarity judgments. U.S. Army Technical Memorandum, Human Engineering Lab., Aberdeen Res. and Development Centre, Maryland, 1972, No. 13-72.

- FENKER, R.M., AND BROWN, D.R. Pattern perception, conceptual spaces, and dimensional limitations on information processing. Multivariate Behav. Res., 1969, 4, 257-271.
- FIELD, J. Feedback modification of human similarity judgments. Project Report, M.A., Dept. Psychol., Univ. Canterbury, 1974.
- FILLENBAUM, S. Problems in the use of similarity data for the study of semantic structures. Rep. L.L. Thurstone Psychometric Lab., Univ. N. Carolina, 1973, No. 114.
- FISCHER, W., AND MICKO, H.C. More about metrics of subjective spaces and attention distributions. J. Math. Psychol., 1972, 9, 36-54.
- FRANZÉN, O., NORDMARK, J., AND SJÖBERG, L. A study of pitch. Göteborg Psychol. Rep., 1972, 2, No. 12.
- FREEDLE, R. A stimulus similarity scale for temporal measures of attention in infants and children. Developmental Psychol., 1971, 4, 240-247.
- GALANTER, E.H. An axiomatic and experimental study of sensory order and measure. Psychol. Rev., 1956, 63, 16-28.
- GIBSON, E.J. Learning to read. Science, 1965, 148, 1066-1072.

GLESER, G.C. Quantifying similarity between people.

In Katz, M.M., Cole, J.O., and Barton, W.E. (Eds.)
Role and methodology of classification in psychiatry
and psychopathology. U.S. National Institute of
 Mental Health, 1968.

GOLDMEIER, E. Similarity in visually perceived forms.

Psychological Issues, 1972, 8, No. 1, monograph 29.

(The first six chapters of this originally appeared
 as: Goldmeier, E. "Über Ähnlichkeit bei gesehenen
 Figuren. Psychol. Forsch., 1936, 21, 146-208).

GOODALL, D.W. A new similarity index based on probability.

Biometrics, 1966, 22, 882-907.

GOUDE, G. A quantitative model for similarity experience.

Rep. Psychol. Lab., Univ. Uppsala, 1966, No. 39.

GOWER, J.C. Some distance properties of latent root and
 vector methods used in multivariate analysis.

Biometrika, 1966, 53, 325-338.

GOWER, J.C. A general coefficient of similarity and some
 of its properties. Unpubl. ms., Rothamsted Exp.
 Station, England, 1967.

GRAHAM, C.H. (Ed.) Vision and visual perception. New York:
 Wiley, 1965.

GREEN, P.E., AND CARMONE, F.J. The effect of task on
 intra-individual differences in similarities

judgments. Multivariate Behav. Res., 1971, 6, 433-450.

GREGORY, R.L. Eye and brain - the psychology of seeing.

London: World University Library, 1966.

- GREGSON, R.A.M. Theoretical and empirical multidimensional scalings of taste mixture matchings. Brit. J. Math. Statist. Psychol., 1966, 19, 59-75.
- GREGSON, R.A.M. Simulating perceived similarities between taste mixtures having mutually interacting components. Brit. J. Math. Statist. Psychol., 1968, 21, 117-130.
- GREGSON, R.A.M. Quadratic similarities. Brit. J. Math. Statist. Psychol., 1970, 23, 53-68.
- GREGSON, R.A.M. Quadratic similarities and a collapse-points condition. J. Math. Psychol., 1972, 9, 368-388.
- GREGSON, R.A.M. Similarity judgments modified by feedback. Acta Psychologica, 1974, 38, 117-129.
- GREGSON, R.A.M. Psychometrics of similarity. New York: Academic Press, 1975.
- GUILFORD, J.P. Psychometric methods. 2nd. ed. New York: McGraw-Hill, 1954.
- GUILFORD, J.P. AND DINGMAN, H.F. A modification of the method of equal-appearing intervals. Amer. J. Psychol., 1955, 68, 450-454.
- GUTTMAN, L. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. Psychometrika, 1968, 33, 469-506.
- HALMOS, P.R. Measure theory. New York: D. Van Nostrand Reinhold, 1950.

HELM, C.E., MESSICK, S., AND TUCKER, L.R. Psychological models for relating discrimination and magnitude estimation scales. Psychol. Rev., 1961, 68, 167-177.

HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. Psychol. Rev., 1948, 55, 297-313.

HÖIJER, B. Mechanisms of cognitive judgments.
Rep. Dept. Psychol., Univ. Uppsala, 1969, No. 67 (a).

HÖIJER, B. On testing metric and nonmetric models of multidimensional similarity. Rep. Dept. Psychol., Univ. Uppsala, 1969, No. 68 (b).

HÖIJER, B. On the consistency of similarity judgments.
Rep. Dept. Psychol., Univ. Uppsala, 1970, No. 77 (a).

HÖIJER, B. Isosimilarity contours of twodimensional cognitive judgments. Rep. Dept. Psychol., Univ. Uppsala, 1970, No. 79 (b).

HYMAN, R., AND WELL, A. Judgments of similarity and spatial models. Perception and Psychophysics, 1967, 2, 233-248.

HYMAN, R., AND WELL, A. Perceptual separability and spatial models. Perception and Psychophysics, 1968, 3, 161-165.

INDOW, T. Colour atlases and colour scaling. In Association Internationale de la Couleur, Color '73: Proceedings of the 2nd Congress. London: Adam Hilger, 1974, p. 137-152 (a).

- INDOW, T. Scaling of saturation and hue shift: Summary of results and implications. In Moskowitz, H.R., et.al (Ed.) Sensation and Measurement. Dordrecht: Reidel, 1974 (b).
- JACCARD, P. The distribution of the flora in the Alpine zone. New Phytologist, 1912, 11, 37-50.
- JOHNSON, S.C. Hierarchical clustering schemes. Psychometrika, 1967, 32, 241-254.
- JUDD, D.B. Basic correlates of the visual system. In Stevens, S.S. (Ed.) Handbook of Experimental Psychology. New York: Wiley, 1960.
- JUNGE, K. The category scale equation. Scand. J. Psychol., 1960, 1, 112-114.
- KELLEY, J.L. General topology. New York: Van Nostrand, 1969.
- KOOPMAN, R.F., AND COOPER, M. Some problems with Minkowski distance models in multidimensional scaling. Presented to the Amer. Psychometric Soc., Stanford, March, 1974.
- KRANTZ, D.H. Rational distance functions for multidimensional scaling. J. Math. Psychol., 1967, 4, 226-245.
- KRANTZ, D.H., AND TVERSKY, A. Similarity of rectangles: An analysis of subjective dimensions. J. Math. Psychol., 1975, 12, 4-34.

KRUSKAL, J.B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis.

Psychometrika, 1964, 29, 1-28 (a).

KRUSKAL, J.B. Nonmetric multidimensional scaling: a numerical method. Psychometrika, 1964, 29, 115-130 (b).

KÜNNAPAS, T., AND KÜNNAPAS, U. On the mechanism of subjective similarity for unidimensional continua. Rep. Psychol. Lab., Univ. Stockholm, 1971, No. 342.

KÜNNAPAS, T., AND KÜNNAPAS, U. On the relation between similarity and ratio estimates. Rep. Psychol., Lab., Univ. Stockholm, 1973, No. 382.

KÜNNAPAS, T., AND KÜNNAPAS, U. On the mechanism of subjective similarity for unidimensional continua. Amer. J. Psychol., 1974, 87, 215-222.

LANCE, G.N., AND WILLIAMS, W.T. Computer programs for hierarchical polythetic classification ("similarity analyses"). Computer J., 1966, 9, 60-64.

LANCE, G.N., AND WILLIAMS, W.T. Mixed-data classificatory programs. I. Agglomerative systems. Austral. Computer J., 1967, 1, 15-20.

LANDAHL, H.D. Neural mechanisms for the concepts of difference and similarity. Bull. Math. Biophysics, 1945, 7, 83-88.

LEVY, R.M., AND KAUFMAN, H.M. Sets and subsets in the identification of multidimensional stimuli. Psychol. Rev., 1973, 80, 139-148.

- LOCKHEAD, G.R. Identification and the form of multidimensional discrimination space. J. Exp. Psychol., 1970, 85, 1-10.
- LUCE, R.D. On the possible psychophysical laws. Psychol. Rev., 1959, 66, 81-95 (a).
- LUCE, R.D. Individual choice behaviour. New York: Wiley, 1959 (b).
- LUCE, R.D. A choice theory analysis of similarity judgments. Psychometrika, 1961, 26, 151-163.
- LUCE, R.D. Detection and recognition. In Luce, R.D., Bush, R.R., and Galanter, E.H. (Eds.) Handbook of Mathematical Psychology. New York: Wiley, 1963, Vol. I.
- LUCE, R.D., BUSH, R.R., AND GALANTER, E.H. (Eds.) Handbook of Mathematical Psychology. New York: Wiley, 1963, Vol. I.
- LUCE, R.D., AND GALANTER, E.H. Discrimination. In Luce, R.D. et.al. (Eds.) Handbook of Mathematical Psychology. New York: Wiley, 1963, Vol. I (a).
- LUCE, R.D., AND GALANTER, E.H. Psychophysical scaling. In Luce, R.D. et. al. (Eds.) Handbook of Mathematical Psychology. New York: Wiley, 1963, Vol. I (b).
- LUCE, R.D., AND GREEN, D.M. A neural timing theory for response times and the psychophysics of intensity. Psychol. Rev., 1972, 79, 14-57.

LUND, T. Multidimensional scaling of political parties - a methodological study. Scand. J. Psychol., 1974, 15, 108-118.

LUND, T. An alternative content method for multidimensional scaling. Multivariate Behav. Res., 1975, 10, 181-191.

LUNEBERG, R.K. The metric of binocular visual space. J. Opt. Soc. Amer., 1950, 40, 627-642.

MACH, E. The analysis of sensations. London: Open Court, 1914.

MARKS, L.E. Sensory processes - the new psychophysics. New York: Academic Press, 1974.

MICHELS, W.C., AND HELSON, H. A reformulation of the Fechner Law in terms of Adaptation-Level applied to rating-scale data. Amer. J. Psychol., 1949, 62, 355-368.

MICKO, H.C. A "halo"-model for multidimensional ratio scaling. Psychometrika, 1970, 35, 199-227.

MICKO, H.C., AND FISCHER, W. The metric of multidimensional psychological spaces as a function of the differential attention to subjective attributes. J. Math. Psychol., 1970, 7, 118-143.

MITCHELL, M.J. Investigations of olfactory similarity scaling. Thesis, Ph.D., Univ. Of Canterbury, 1971.

- NASH, M.C. An experimental test of the Michels-Helson theory of judgment. Amer. J. Psychol., 1950, 63, 214-220.
- PARDUCCI, A. Category judgment: A range-frequency model. Psychol. Rev., 1965, 72, 407-418.
- PFANZAGL, J. Theory of measurement. Würzburg/Vienna: Physica-Verlag, 1968.
- POSTMAN, L., AND MILLER, G.A. Anchoring of temporal judgments. Amer. J. Psychol., 1945, 58, 43-53.
- REED, S.K. Pattern recognition and categorization. Cognitive Psychol., 1972, 3, 382-407.
- RESTLE, F. A metric and an ordering on sets. Psychometrika, 1959, 24, 207-219.
- RESTLE, F. Psychology of judgment and choice: A theoretical essay. New York: Wiley, 1961.
- RICHARDSON, M.W. Multidimensional psychophysics. Psychol. Bull., 1938, 35, 659-660.
- ROSENBERG, W.J. On the validity of Minkowski metrics: A critical note on Roskam's scaling of Eisler's similarities data. Ned. T. Psychol., 1974, 29, 229-233.
- ROSKAM, E.E. Metric analysis of ordinal data in psychology. Voorschoten, the Netherlands, 1968.

ROSKAM, E.E. Data Theory and metrical analysis, Part I.

Ned. T. Psychol., 1970, 25, 15-54 (a).

ROSKAM, E.E. Data Theory and algorithms for nonmetric

scaling, Part II. Ned. T. Psychol., 1970, 25,

66-82 (b).

ROSKAM, E.E. Multidimensional scaling by metric

transformation of data. Ned. T. Psychol., 1972, 27,

486-508.

RUMELHART, D.L., AND GREENO, J.G. Similarity between

stimuli: An experimental test of the Luce and

Restle choice models. J. Math. Psychol., 1971, 8,

370-381.

SHEPARD, R.N. Stimulus and response generalisation: A

stochastic model relating generalisation to distance

in psychological space. Psychometrika, 1957, 22,

325-345.

SHEPARD, R.N. Stimulus and response generalisation: Tests

of a model relating generalisation to distance in

psychological space. J. Exp. Psychol., 1958, 55,

509-523 (a).

SHEPARD, R.N. Stimulus and response generalisation:

Deduction of the generalisation gradient from a

trace model. Psychol. Rev., 1958, 65, 242-256 (b).

SHEPARD, R.N. Similarity of stimuli and metric properties of behavioural data. In Gulliksen, H., and Messick, S. (Eds.) Psychological scaling: Theory and applications. New York: Wiley, 1960.

SHEPARD, R.N. The analysis of proximities: multidimensional scaling with an unknown distance function, Part I. Psychometrika, 1962, 27, 125-140 (a).

SHEPARD, R.N. The analysis of proximities. Multidimensional scaling with an unknown distance function, Part II. Psychometrika, 1962, 27, 219-246. (b)

SHEPARD, R.N. Attention and the metric structure of the stimulus space. J. Math. Psychol., 1964, 1, 54-87.

SHEPARD, R.N., ROMNEY, A.K., AND NERLOVE, S.B. Multidimensional scaling: Theory and applications in the behavioural sciences. Vol I - Theory. Vol. II - Applications. New York: Seminar Press, 1972.

SJÖBERG, L. Models of similarity and intensity. Göteborg Psychol. Rep., 1973, 3, No. 10.

SPENCE, I., AND GRAEF, J. The determination of the underlying dimensionality of an empirically obtained matrix of proximities. Multivariate Behav. Res., 1974, 9, 331-341.

STEINHEISER, F.H. Individual preference scales within a multidimensional "similarities" space. J. Exp. Psychol., 1970, 86, 325-327.

STEVENS, S.S. On the psychophysical law. Psychol. Rev., 1957, 64, 153-181.

STEVENS, S.S. Adaptation-level vs the relativity of judgment. Amer. J. Psychol., 1958, 71, 633-647.

STEVENS, S.S. Mathematics, measurement and psychophysics. In Stevens, S.S. (Ed.) Handbook of Experimental Psychology, New York: Wiley, 1960 (a).

STEVENS, S.S. Ratio scales, partition scales and confusion scales. In Gulliksen, H., and Messick, S. (Eds.) Psychological scaling: Theory and applications. New York: Wiley, 1960 (b).

STEVENS, S.S. A metric for the social consensus. Science, 1966, 151, 530-541.

STEVENS, S.S., AND GALANTER, E.H. Ratio scales and category scales for a dozen perceptual continua. J. Exp. Psychol., 1957, 54, 377-411.

TERNUS, J. The problem of phenomenal identity. In Ellis, W.D. (Ed.) A sourcebook of Gestalt psychology. New York: Harcourt, Brace, 1938.

THURSTONE, L.L. Multiple-factor analysis: A development and expansion of "The Vectors of the Mind". Chicago: Univ. Chicago, 1947.

TORGERSON, W.S. Theory and methods of scaling. New York: Wiley, 1958.

- TORGERSON, W.S. Quantitative judgment scales. In
Gulliksen, H., and Messick, S. (Eds.) Psychological
scaling: Theory and applications. New York: Wiley, 1960.
- TORGERSON, W.S. Multidimensional scaling of similarity.
Psychometrika, 1965, 30, 379-393.
- TVERSKY, A. Elimination by aspects: A theory of choice.
Psychol. Rev., 1972, 79, 281-299.
- TVERSKY, A., AND KRANTZ, D.H. Similarity of schematic
faces: A test of interdimensional additivity.
Perception and Psychophysics, 1969, 5, 124-128.
- TVERSKY, A., AND KRANTZ, D.H. The dimensional representation
and the metric structure of similarity data. J.
Math. Psychol., 1970, 7, 572-596.
- VEGELIUS, J. Correlation coefficients as scalar products
in Euclidean spaces. Rep. Dept. Psychol., Univ.
Uppsala, 1973, No. 145.
- WAERN, Y. A model for multidimensional similarity.
Rep. Psychol. Lab., Univ. Stockholm, 1968, No.256 (a).
- WAERN, Y. Multidimensional scaling with a priori
dimensions. Rep. Psychol. Lab., Univ. Stockholm, 1968,
No. 270 (b).
- WAERN, Y. Multidimensional similarity: an analysis of
'analyzability'. Rep. Psychol. Lab., Univ. Stockholm,
1969, No. 280.

- WAERN, Y. Preference and perception: A study on individual subjects of the unfolding model. Rep. Psychol. Lab., Univ. Stockholm, 1970, No. 309 (a).
- WAERN, Y. A comparison of some models for multidimensional similarity. Rep. Psychol. Lab., Univ. Stockholm, 1970, No. 312 (b).
- WAERN, Y. A model for multidimensional similarity. Percept. Mot. Skills, 1971, 33, 15-25. (a).
- WAERN, Y. Structure in similarity matrices - a graphic approach. Scand. J. Psychol., 1971, 12, 5-16(b).
- WAERN, Y. Similarity estimates and cognitive structure. Studies in some models and methods. Rep. Psychol. Lab., Univ. Stockholm, 1971, supplement 9. (c)
- WAGENAAR, W.A. Stevens vs Fechner: A plea for dismissal of the case. Acta Psychologica, 1975, 39, 225-235.
- WALLACH, M.A. On psychological similarity. Psychol. Rev., 1958, 65, 103-116.
- WENDER, K. Interpretation of Minkowski metrics in ordinal multidimensional scaling. Presented to the 2nd Annual Meeting of the Gruppe Math. Psychol., Marburg/Lahn, 1968.
- WENDER, K. A test of independence of dimensions in multidimensional scaling. Perception and Psychophysics, 1971, 10, 30-32.

WERTHEIMER, M. Laws of organisation in perceptual forms.

In Ellis, W.D. (Ed.) A sourcebook of Gestalt Psychology. New York: Harcourt Brace, 1938.

YOUNG, F.W. Nonmetric multidimensional scaling:

Development of an index of metric determinacy.

Rep. L.L. Thurstone Psychometric Lab., Univ. N. Carolina, 1968, No. 68.

YOUNG, F.W. A model for polynomial conjoint analysis

algorithms. In Shepard, R.N., Romney, A.K., and

Nerlove, S.B. (Eds.) Multidimensional scaling: Theory and applications in the behavioural sciences.

New York: Seminar Press, 1972, Vol. I.

YOUNG, F.W. Conjoint Scaling. Rep. L.L. Thurstone Psycho-

metric Lab., Univ. N. Carolina, 1973, No. 118.

YOUNG, G., AND HOUSEHOLDER, A.S. Discussion of a set of points in terms of their mutual distances.

Psychometrika, 1938, 3, 19-22.

ZAGORSKI, M.A. A topological test of metric models of

stimulus similarity. Presented to seventh Annual

Math. Psychol. Meeting, Ann Arbor, 1974.

APPENDIX A

PROOFS USED IN CHAPTER 4.

1. $D_{1\beta}$ obeys the triangle inequality strictly on straight lines, for $\beta \geq 1$ (see section 4.1.1.1):

We shall use the notation

$$\|x\| = \left(\sum_{i=1}^n |x_i|^\beta \right)^{1/\beta} \quad \forall x \in \mathbb{R}^n$$

$$\text{so that } D_{1\beta}(x, y) = \frac{\|x - y\|}{\|x\| + \|y\|} \quad \forall x, y \in \mathbb{R}^n.$$

Let y be between x and z on a straight line; then there exists an $\alpha \in \mathbb{R}^+$, $0 \leq \alpha \leq 1$, such that

$$y_i = \alpha x_i + (1-\alpha)z_i \quad \forall i \in \{1, 2, \dots, n\}$$

(i.e. $y = \alpha x + (1-\alpha)z$).

Thus we have that

$$\begin{aligned} & D_{1\beta}(x, y) + D_{1\beta}(y, z) \\ &= \frac{(1-\alpha) \|x - z\|}{\|\alpha x + (1-\alpha)z\| + \|x\|} + \frac{\alpha \|x - z\|}{\|\alpha x + (1-\alpha)z\| + \|z\|} \\ &= \|x - z\| \frac{\|\alpha x + (1-\alpha)z\| + \alpha \|x\| + (1-\alpha) \|z\|}{(\|\alpha x + (1-\alpha)z\| + \|z\|)(\|\alpha x + (1-\alpha)z\| + \|x\|)}. \end{aligned}$$

Now, by the Minkowski Inequality (since we have assumed that $\beta \geq 1$),

$$\| \alpha x + (1-\alpha)z \| \leq \alpha \| x \| + (1-\alpha) \| z \| \quad (\text{A.1})$$

$$\text{so } \| \alpha x + (1-\alpha)z \| \leq \max\{ \| x \| , \| z \| \}. \quad (\text{A.2})$$

(a) Suppose that

$$\| \alpha x + (1-\alpha)z \| \geq \min\{ \| x \| , \| z \| \} \quad (\text{A.3}).$$

From (A.1) we have

$$\begin{aligned} D_{1\beta}(x,y) + D_{1\beta}(y,z) &\geq \\ &\frac{2 \| x-z \| \| \alpha x + (1-\alpha)z \|}{\| \alpha x + (1-\alpha)z \| (\| x \| + \| z \|) + \| \alpha x + (1-\alpha)z \|^2 + \| x \| \| z \|} \end{aligned}$$

But

$$\begin{aligned} &\| \alpha x + (1-\alpha)z \|^2 + \| x \| \| z \| \\ = & (\| \alpha x + (1-\alpha)z \| - \| x \|) (\| \alpha x + (1-\alpha)z \| - \| z \|) \\ &+ \| \alpha x + (1-\alpha)z \| (\| x \| + \| z \|) \\ \leq & \| \alpha x + (1-\alpha)z \| (\| x \| + \| z \|) \\ &\text{since, by (A.3) and (A.2)} \\ &(\| \alpha x + (1-\alpha)z \| - \| x \|) (\| \alpha x + (1-\alpha)z \| - \| z \|) \leq 0. \end{aligned}$$

Therefore

$$\begin{aligned} D_{1\beta}(x,y) + D_{1\beta}(y,z) &\geq \\ \| x-z \| \frac{2 \| \alpha x + (1-\alpha)z \|}{2 \| \alpha x + (1-\alpha)z \| (\| x \| + \| z \|)} &= \frac{\| x-z \|}{\| x \| + \| z \|} \\ &= D_{1\beta}(x,z) \end{aligned}$$

(b) Suppose that

$$\| \alpha x + (1-\alpha)z \| \leq \min\{ \| x \| , \| z \| \} \quad (\text{A.4}).$$

But we have, in either case, that

$$\alpha \|x\| + (1-\alpha) \|z\| \geq \min\{\|x\|, \|z\|\}$$

which implies that

$$\begin{aligned} D_{1\beta}(x,y) + D_{1\beta}(y,z) &\geq \\ &\frac{\|x-z\|}{\|\alpha x + (1-\alpha)z\| + \max\{\|x\|, \|z\|\}} \\ &\geq \frac{\|x-z\|}{\min(\|x\|, \|z\|) + \max\{\|x\|, \|z\|\}} \\ &\quad \text{by (A.4)} \\ &= \frac{\|x-z\|}{\|x\| + \|z\|} = D_{1\beta}(x,z). \end{aligned}$$

But (A.3) and (A.4) are exhaustive of all the relevant possible cases, so the proof of the triangle inequality is complete.

To have equality in case (a), we must have equality both in (A.1) and in either (A.2) or (A.3). Equality in (A.1) occurs only if

$$\exists \mu \in \mathbb{R} \ni x_i = \mu z_i \quad i = 1, 2, \dots, n. \quad (\text{A.5})$$

which implies that $\|x\| = \mu \|z\|$. But equality will then occur in (A.2) or (A.3) only if

$$\|x\| = \|z\| \quad (\text{A.6}):$$

i.e., if $\mu = 1$ - which means that $x = z$, a contradiction.

Hence there must be strict inequality in case (a).

In case (b), equality will occur only if (A.4) has equality and

$$\alpha \|x\| + (1-\alpha) \|z\| = \min(\|x\|, \|z\|).$$

But the two equations give

$$\|\alpha x + (1-\alpha)z\| = \min(\|x\|, \|z\|) = \alpha \|x\| + (1-\alpha) \|z\|$$

so again we must have (A.5) and (A.6): a contradiction.

Thus there is strict inequality in all cases.

2. If $x, y, a(i) \in \mathbb{R}^n$, where $i \in \{1, 2, \dots, n\}$ and $a(i)_i = 0$,

then

$$D_{11}(x_i + a(i), y_i + a(i)) \leq D_{11}(x, y) \quad (\text{A.5})$$

if

$$\sum_{j \neq i} |a(i)_j| \geq \min\left\{ \sum_{j \neq i} |x_j|, \sum_{j \neq i} |y_j| \right\} \quad (\text{A.6})$$

(see equations (4.4) and (4.7) in section 4.1.1.1):

Let $A = \sum_{j \neq i} |a(i)_j|$. Then (A.5) is equivalent to

$$\frac{|x_i - y_i|}{2A + |x_i| + |y_i|} \leq \frac{\sum_{j=1}^n |x_j - y_j|}{\sum_{j=1}^n |x_j| + \sum_{j=1}^n |y_j|}$$

or, rearranging,

$$A \geq \frac{1}{2} \left[\frac{\sum_{j=1}^n |x_j| + \sum_{j=1}^n |y_j|}{\sum_{j=1}^n |x_j - y_j|} |x_1 - y_1| - (|x_1| + |y_1|) \right] \quad (A.7)$$

Calling the right hand side of this inequality

R, we have

$$R = \frac{1}{2} \left[\frac{|x_1 - y_1| \sum_{j=1}^n (|x_j| + |y_j|) - (|x_1| + |y_1|) \sum_{j=1}^n |x_j - y_j|}{\sum_{j=1}^n |x_j - y_j|} \right]$$

$$= \frac{1}{2} \left[\frac{|x_1 - y_1| \sum_{j \neq 1}^n (|x_j| + |y_j|) - (|x_1| + |y_1|) \sum_{j \neq 1}^n |x_j - y_j|}{\sum_{j=1}^n |x_j - y_j|} \right]$$

so, since $|x_1 - y_1| \leq |x_1| + |y_1|$

$$R \leq \frac{|x_1 - y_1|}{\sum_{j=1}^n |x_j - y_j|} \frac{1}{2} \left[\sum_{j \neq 1}^n (|x_j| + |y_j| - |x_j - y_j|) \right]$$

$$\leq \frac{1}{2} \sum_{j \neq 1}^n (|x_j| + |y_j| - |x_j - y_j|)$$

$$\leq \frac{1}{2} \sum_{j \neq 1}^n (|x_j| + |y_j| - ||x_j - y_j||)$$

$$= \sum_{j \neq 1}^n \min\{|x_j|, |y_j|\}$$

$$\leq \min\left\{ \sum_{j \neq 1}^n |x_j|, \sum_{j \neq 1}^n |y_j| \right\}$$

Hence, if $A \geq \min(\sum_{j \neq 1} |x_j|, \sum_{j \neq 1} |y_j|)$, then
 $A \geq R$, which is (A.7) Thus, by the definitions of
 A and R , (A.6) implies (A.5), Q.E.D.

3. $F_2(D_{12})$ is multiplicative on straight lines
 (see section 4.1.1.3):

Recall that F_2 is defined as

$$F_2(\alpha) = \frac{1-\alpha}{1+\alpha} \quad 0 \leq \alpha \leq 1$$

Thus three numbers α, β, γ are multiplicative under F_2
 - say

$$F_2(\alpha) = F_2(\beta)F_2(\gamma) \quad -$$

if and only if

$$\alpha = \frac{\beta + \gamma}{1 + \beta\gamma} \quad .$$

If x, y, z lie on a straight line with y between x and
 z : i.e.

$$y = \delta x + (1-\delta)z \text{ for some } \delta, 0 \leq \delta \leq 1;$$

we therefore need to look at

$$B = \frac{D_{12}(x, y) + D_{12}(y, z)}{1 + D_{12}(x, y) D_{12}(y, z)}$$

and show

$$B = D_{12}(x, z).$$

But, using the notation

$$\|x\| = \left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}} \quad \forall x \in \mathbb{R}^n,$$

we have

$$\begin{aligned}
 B &= \frac{\|x - \delta x - (1-\delta)z\|}{\|x\| + \|\delta x + (1-\delta)z\|} + \frac{\|\delta x + (1-\delta)z - z\|}{\|\delta x + (1-\delta)z\| + \|z\|} \\
 &= \frac{1 + \frac{\|x - \delta x - (1-\delta)z\| \cdot \|\delta x + (1-\delta)z - z\|}{(\|x\| + \|\delta x + (1-\delta)z\|)(\|\delta x + (1-\delta)z\| + \|z\|)}}{\frac{\|x - z\|((1-\delta)(\|z\| + \|\delta x + (1-\delta)z\|) + \delta(\|x\| + \|\delta x + (1-\delta)z\|))}{\|x\| \|z\| + \|\delta x + (1-\delta)z\|(\|x\| + \|z\|) + \|\delta x + (1-\delta)z\|^2 + \delta(1-\delta)\|x - z\|^2}} \\
 &= \frac{\|x - z\|((1-\delta)(\|z\| + \delta\|x\| + \|\delta x + (1-\delta)z\|))}{\delta\|x\| \|z\| + (1-\delta)\|x\| \|z\| + \|\delta x + (1-\delta)z\|(\|x\| + \|z\|) + \delta\|x\|^2 + (1-\delta)\|z\|^2} \\
 &= \frac{\|x - z\|}{\|x\| + \|z\|} \\
 &= D_{12}(x, z) \quad \text{Q.E.D.}
 \end{aligned}$$

4. With points (x, y, z) either on a straight line under D_{12} , or on a straight line through the origin under D_{1g} , if y is between x and z , then

$$g(x, z) < g(x, y)g(y, z) \quad (\text{A.8})$$

where $g = F_1(D)$ (see Table 4.1):

Firstly, note that since $g(x, z) \leq g(x, y), g(y, z) \leq 1$, trivially we have

$$g(x, y) \geq g(x, z)g(y, z)$$

and $g(y, z) \geq g(x, z)g(x, y)$.

Thus the inequality (A.8) is the critical one. If (A.8) holds then we have shown that the similarity triangle inequality holds only trivially on straight lines under D_{12} , and on straight lines through the origin under

$D_{1\beta}$, with $\beta = F_1(D)$.

To prove (A.8), we use the fact, shown in the previous section of this Appendix, and in section 4.2, that in the cases under consideration here,

$$F_2(D(x,z)) = F_2(D(x,y))F_2(D(y,z))$$

which is equivalent to

$$D(x,z) = \frac{D(x,y) + D(y,z)}{1 + D(x,y)D(y,z)} \quad (A.9)$$

since $F_2(\alpha) = (1-\alpha)/(1+\alpha)$, $0 \leq \alpha \leq 1$.

Substituting from (A.9) we see that (A.8) is equivalent to

$$F_1\left(\frac{D(x,y)+D(y,z)}{1+D(x,y)D(y,z)}\right) < F_1(D(x,y))F_1(D(y,z))$$

or, putting $\gamma = D(x,y)$ and $\delta = D(y,z)$,

$$1 - \frac{\gamma+\delta}{1+\gamma\delta} < (1-\gamma)(1-\delta) \quad (A.10).$$

But

$$\begin{aligned} & (1-\gamma)(1-\delta) - \left(1 - \frac{\gamma+\delta}{1+\gamma\delta}\right) \\ &= \frac{\gamma+\delta-\gamma-\delta+\gamma\delta+\gamma\delta(\gamma\delta-\gamma-\delta)}{1+\gamma\delta} = \frac{\gamma\delta(1-\gamma)(1-\delta)}{1+\gamma\delta} \\ &> 0 \text{ since } 0 \leq \gamma, \delta \leq 1 \end{aligned}$$

which proves (A.10) and therefore (A.8).

Q.E.D.

APPENDIX B.PROOF USED IN CHAPTER 5.

If all pairs of consecutive category exemplars of a category scale are in equal ratio, then the category numbers are linear with the logarithms of their respective exemplars (see section 5.1.2):

Consider a category scale of n (>2) categories, x_i being the exemplar of the i th category, with

$$a = x_1, \quad b = x_n \quad (B.1),$$

so a and b are the end-exemplars of the scale. We shall denote the category scale value corresponding to the stimulus with value x as $K(x)$; thus in particular

$$K(x_i) = i \quad i = 1, 2, \dots, n. \quad (B.2)$$

By assumption we have

$$\frac{x_i}{x_{i+1}} = \frac{x_j}{x_{j+1}} \quad \forall i, j \in \{1, 2, \dots, n-1\} \quad (B.3)$$

We shall call any set of stimulus values $\{x_i\}_{i=1}^n$,

with $a = x_1$, $b = x_n$, an n -section of (a, b) if (B.3) is true of the set.

We now show by induction that for any n -section $\{x_i\}_{i=1}^n$ of any (a,b) , if $n \geq 2$, then

$$x_i = a \left(\frac{b}{a}\right)^{\frac{i-1}{n-1}} \quad i=1, \dots, n. \quad (\text{B.4})$$

It is obviously true for $n = 2$. If $n = 3$ then by (B.3), for any (a,b) ,

$$\frac{a}{x_2} = \frac{x_2}{b}$$

$$\text{or } x_2 = \sqrt{ab} = a \left(\frac{b}{a}\right)^{\frac{1}{2}}$$

so it is true for $n = 3$.

Suppose now that (B.4) is true for any (a,b) for $n = k$ ($k \geq 3$).

Let $n = k+1$. Take a $(k+1)$ -section $\{x_i\}_{i=1}^{k+1}$ of an interval (a,b) where $a=x_1$, $b=x_{k+1}$. Then clearly $\{x_i\}_{i=1}^k$ is a k -section of the interval (a, x_k) , so by our last assumption

$$x_i = a \left(\frac{x_k}{a}\right)^{\frac{i-1}{k-1}} \quad i=1, \dots, k. \quad (\text{B.5})$$

Similarly $\{x_i\}_{i=2}^{k+1}$ is a k -section of (x_2, b) ,

$$x_i = x_2 \left(\frac{b}{x_2}\right)^{\frac{i-2}{k-1}} \quad i=2, \dots, k+1 \quad (\text{B.6}).$$

Putting $i=2$ in (B.5) gives

$$x_2 = a \left(\frac{x_k}{a} \right)^{\frac{1}{k-1}}$$

and substituting in (B.6) with $i=k$, we get

$$x_k = a \left(\frac{x_k}{a} \right)^{\frac{1}{k-1}} \left(\frac{b}{a} \left(\frac{a}{x_k} \right)^{\frac{1}{k-1}} \right)^{\frac{k-2}{k-1}}$$

or, solving for x_k , since $k \geq 3$,

$$x_k = a \left(\frac{b}{a} \right)^{\frac{k-1}{k}}$$

Substituting for x_k in (B.5) gives

$$x_i = a \left(\frac{a \left(\frac{b}{a} \right)^{\frac{k-1}{k}}}{a} \right)^{\frac{i-1}{k-1}} \quad i=1, \dots, k$$

or

$$x_i = a \left(\frac{b}{a} \right)^{\frac{i-1}{k}} \quad i=1, \dots, k.$$

Since $x_{k+1} = b$, this shows that (B.4) holds for $n=k+1$.

Since (B.4) holds for $n=3$, it holds for all $n \geq 3$, so we have proved what we set out to with regard to (B.4).

Now solve each equation (B.4) for i :

$$i = (n-1) \frac{\log(x_i/a)}{\log(b/a)} + 1 \quad i=1, \dots, n \quad (\text{B.7})$$

(where the logarithm is to any base).

But by (B.2), for each i , $K(x_i) = i$, so

$$K(x_i) = (n-1) \frac{\log(x_i/a)}{\log(b/a)} + 1 \quad i=1, \dots, n; \quad n \geq 2$$

which is the required result.